LE SCIENZE
edizione italiana di SCIENTIFIC AMERICAN

numero 148
dicembre 1980
anno XIII
volume XXV

Lo strano spettro di SS 433

Lo spostamento delle righe di emissione scoperto nello spettro di questo straordinario oggetto stellare fa pensare che esso espella due sottili getti di materia ad alta velocità in direzioni opposte

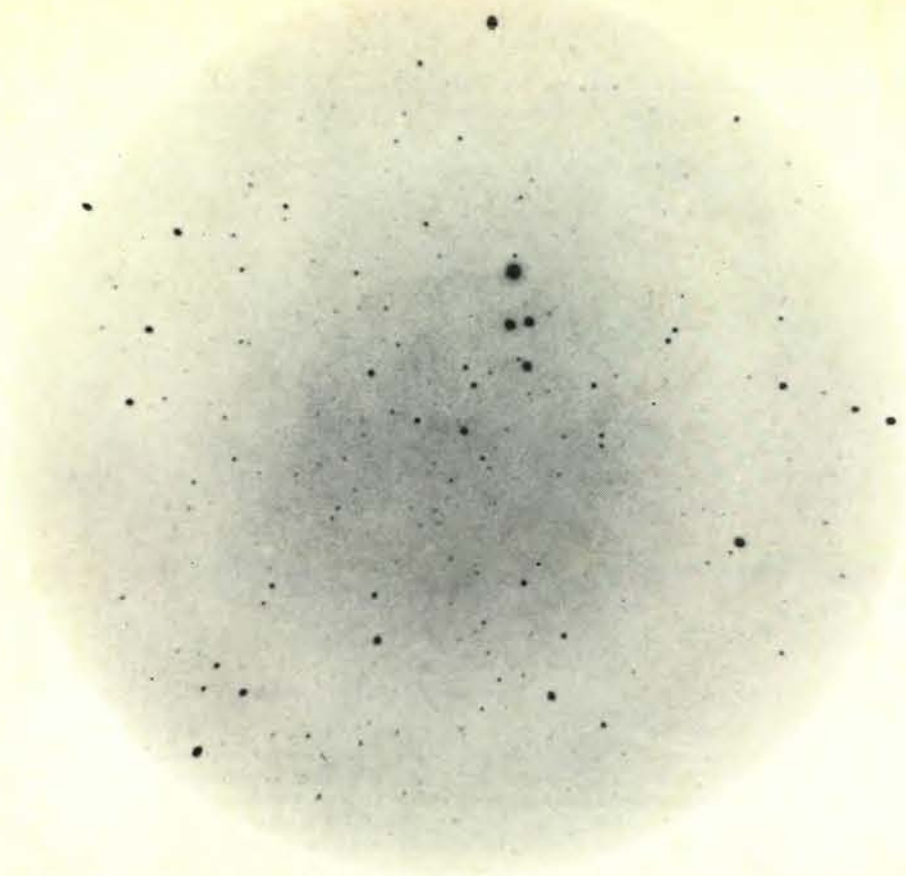
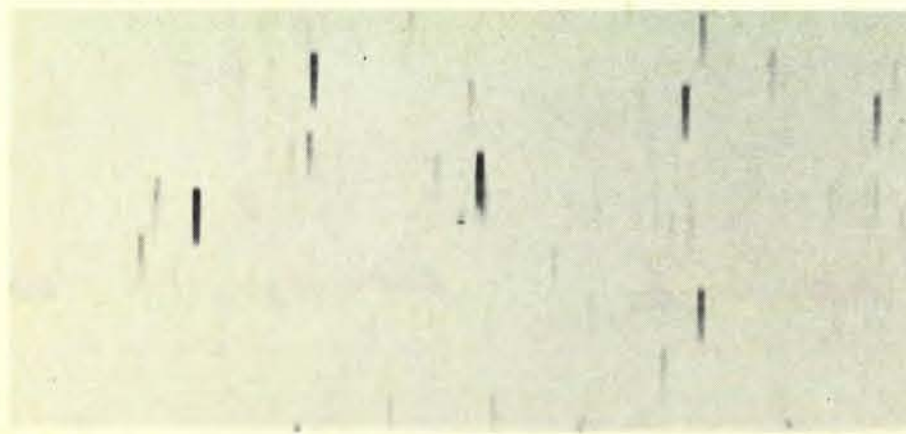
di Bruce Margon

Con 100 miliardi di stelle, nella Via Lattea non resta molto spazio per l'individualità. Virtualmente ogni fenomeno della Galassia ha una certa probabilità di verificarsi più di una volta. Quindi «unico» non è un termine che si possa applicare spesso nell'astronomia stellare. Perfino alcuni oggetti astronomici eccezionalmente insoliti hanno rivelato di possedere compagni simili; per esempio, la Nebulosa del Granchio, il resto spettacolare di un'esplosione di supernova osservata nel 1054, pare abbia due o tre controfigure in altre regioni della Galassia. Tuttavia, negli ultimi due anni si sono accumulate osservazioni di un oggetto stellare che si merita probabilmente l'appellativo di unico. Pare che tale oggetto, noto come SS 433, espella due sottili getti di materia in direzioni opposte con velocità incredibilmente elevate. Non è mai stata osservata alcuna stella il cui comportamento presenti delle caratteristiche analoghe.

SS 433 è un oggetto affascinante per molti motivi, non ultimo il fatto che è stata scoperta indipendentemente parecchie volte negli ultimi vent'anni, pur se i suoi scopritori non sono mai riusciti a evidenziare le sue caratteristiche altamente singolari. Tale oggetto emette una figura di radiazione eccezionale in molte regioni differenti dello spettro elettromagnetico (a lunghezze d'onda che corrispondono a luce visibile, radioonde e raggi X), e ciascuna di tali peculiarità è stata osservata individualmente nel corso degli anni. È stato però solo di recente che ha cominciato a delinearsi un quadro comprensibile di SS 433. È utile rifare daccapo la storia delle osservazioni, per capire quanto si sono avvicinati i primi ricercatori alla scoperta delle singolari proprietà di tale oggetto.

I primi risultati interessanti furono ricavati all'inizio degli anni sessanta quando alcuni astronomi della Case Western Reserve University eseguirono una ricerca sistematica mirante a scoprire certi tipi di stelle deboli nei pressi del piano centrale della nostra galassia. La particolare tecnica telescopica da essi usata comprendeva un prisma obiettivo, un elemento ottico disperdente che allarga le immagine stellari su una lastra fotografica in piccolissime strisce che rappresentano lo spettro visibile della luce stellare. Diversamente dalla tecnica più convenzionale della spettroscopia astronomica, nella quale la luce proveniente da una singola stella viene fatta passare attraverso una stretta fenditura vicina al fuoco del telescopio e successivamente dispersa su una vasta regione della lastra fotografica, la tecnica del prisma obiettivo separa solo di poco in lunghezze d'onda la luce stellare. Essa lo fa simultaneamente per ogni immagine stellare che sia osservabile nel campo visivo, anziché per una stella alla volta. In questo modo si possono ricavare per grandi numeri di stelle alcuni spettrogrammi molto rozzi, atti più che altro alla classificazione dei diversi tipi di oggetti ricorrendo a un modesto impiego del prezioso tempo del telescopio.

L'esplorazione di Case è stata progettata esplicitamente allo scopo di scoprire nuove stelle con righe di emissione nello spettro. L'esistenza di tali righe è la base di una versatile tecnica diagnostica comune a molti campi della fisica e della chimica. Quando gli elettroni legati agli atomi in qualsiasi gas diffuso vengono eccitati in stati energetici «orbitali» più elevati, o per collisioni con altre particelle o per interazioni radiative, la loro successiva diseccitazione crea nuovi fotoni, o quanti di luce, a certe lunghezze d'onda discrete che sono determinate dalla differenza di energia tra gli stati orbitali e quindi dalla struttura elettronica del particolare elemento. La comparsa di tali righe (così chiamate perché si manifestano come un annerimento rettilineo a una particolare lunghezza d'onda sulla lastra fotografica esposta) in uno spettrogramma può quindi servire come una verifica altamente specifica delle condizioni fisiche esistenti nel gas emittente. La esatta lunghezza d'onda di una tale riga, per esempio, è un indicatore indiscutibile dell'elemento chimico originale che emette la luce, dal momento che essa immagazzina in memoria la struttura elettronica dell'atomo.

Studi col prisma obiettivo quali quelli che sono stati eseguiti dai ricercatori di

SS 433 è del tutto irrilevante nelle tradizionali fotografie astronomiche come questa ripresa da Eugene A. Harlan con il telescopio rifrattore da 36 pollici dell'Osservatorio Lick dell'Università di California. L'immagine di SS 433 è il punto nero che si trova esattamente al centro di questo negativo fotografico circolare, che si estende a un campo visivo di circa un decimo di grado nella costellazione dell'Aquila. Classificata in termini della sua luminosità apparente come un oggetto di 14ª grandezza, SS 433 è 1000 volte troppo debole per esser vista a occhio nudo. L'evidente assenza di qualsiasi caratteristica visibile in grado di evidenziare l'oggetto dalle molte stelle vicine in questa regione molto popolata della Via Lattea ci spiega perché esso è stato ignorato a lungo.

Una prima spiegazione dell'insolita natura di SS 433 venne scoperta su questa lastra spettroscopica a campo visivo relativamente ampio ottenuta più di 20 anni fa nel corso di una particolare esplorazione stellare eseguita da astronomi del Warner and Swasey Observatory della Case Western Reserve University. La lastra era stata esposta con la tecnica del prisma obiettivo, che aveva l'effetto di allargare le immagine stellari in corte tracce che rappresentavano lo spettro visibile della luce stellare. In questo caso l'esposizione era limitata mediante l'uso di opportuni filtri alla regione rossa dello spettro. Nella parte del cielo coperta da questa particolare lastra (una superficie larga circa mezzo grado, pari a circa il diametro della Luna piena) solo una immagine spettrale appare come una riga anziché come una macchia allargata; è il debole tratto nei pressi del centro della lastra. La maggior parte della luce rossa proveniente da questo oggetto è concentrata nella intensissima riga di emissione dell'idrogeno alfa, anziché essere distribuita con continuità su tutte le lunghezze d'onda di questo intervallo, come accade nel caso delle stelle vicine. Nel 1977 C. Bruce Stephenson e Nicholas Sanduleak della Case Western Reserve pubblicarono un elenco di tali oggetti a intensa riga di emissione in questa regione del cielo. L'oggetto al centro della lastra occupava la 433ª posizione dell'elenco di Stephenson-Sanduleak e venne chiamato SS 433.

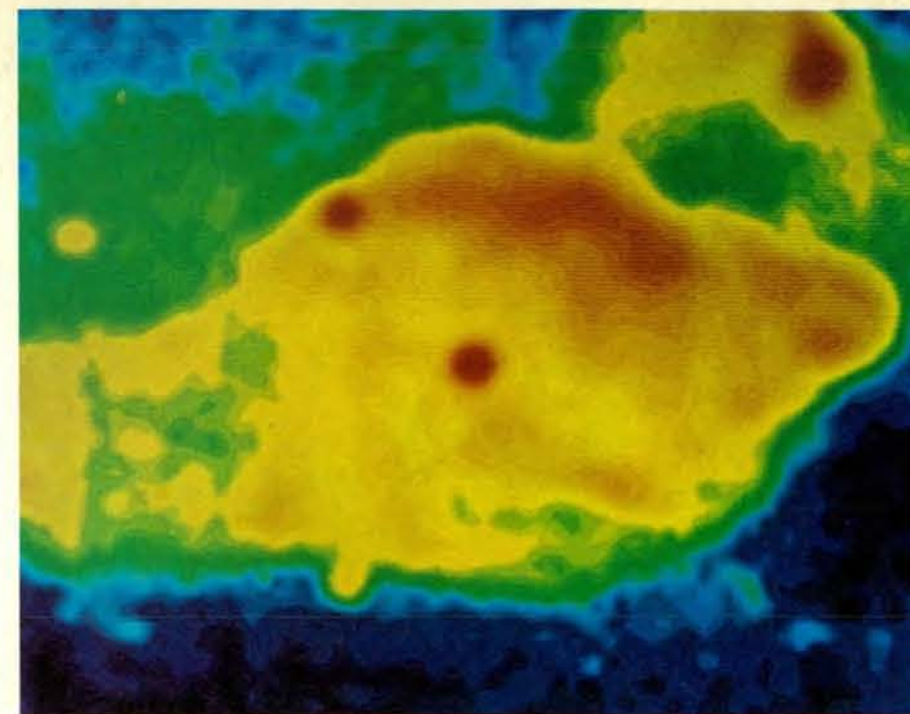
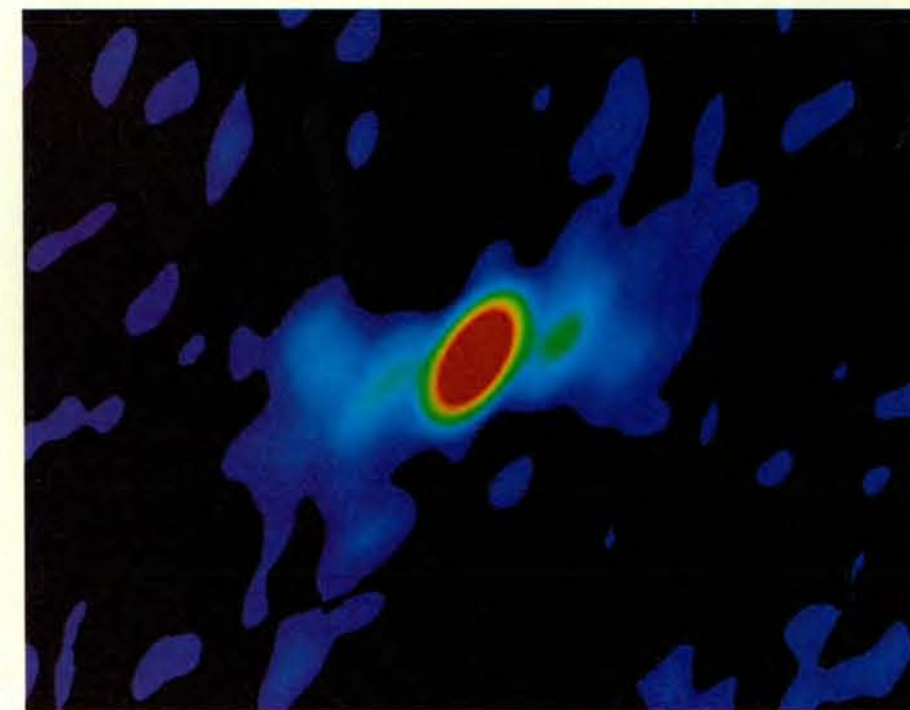
Case hanno dimostrato che forse il 10 per cento delle stelle produce righe di emissione caratteristiche di un gas caldo eccitato; si pensa che tale gas esista in uno strato diffuso esterno che circonda una stella sotto altri aspetti normale. Esistono molte possibili ragioni di questa situazione anomala, variabili da una stella all'altra. Per esempio si trovano spesso righe di emissione associate a stelle molto giovani e a stelle molto vecchie. Talvolta esse vengono rivelate anche in stelle stabili di media età.

Una delle lastre esposte dai ricercatori di Case venne puntata sulla costellazione dell'Aquila nel centro della Via Lattea; sulla lastra comparvero molti oggetti con intense righe di emissione. Nel 1977 Bruce Stephenson e Nicholas Sanduleak pubblicarono un elenco degli oggetti con righe di emissione presenti in tale regione; al posto 433 del loro elenco c'era la stella oggi nota come SS 433. Stephenson e Sanduleak non avevano alcun motivo per considerare SS 433 particolarmente diversa da altre stelle con righe di emissione identificate nella loro esplorazione; a tale scopo sarebbero stati necessari dati spettroscopici più dettagliati di quelli che si potevano ottenere con la tecnica del prisma obiettivo. In effetti, uno dei principali obiettivi di esplorazioni come quella di Case è quello di stimolare future osservazioni dettagliate di oggetti dall'aspetto interessante. Come risultò in seguito, Stephenson e Sanduleak da questo punto di vista ottennero con SS 433 risultati ben al di là delle più ottimistiche previsioni.

Con una luminosità apparente equivalente a quella di una stella di 14ª grandezza, SS 433 è più di 1000 volte più debole per poter essere vista a occhio nudo. È però un facile bersaglio per le osservazioni spettroscopiche, perfino con un telescopio piuttosto piccolo. Le successive osservazioni si possono benissimo far risalire fino al 1975, quando Lawrence Krumenaker, un altro astronomo di Case, pubblicò un breve articolo che comparve prima della pubblicazione del catalogo di Stephenson-Sanduleak. L'articolo citava un piccolo sottogruppo di stelle con righe di emissione estratto dall'intera esplorazione, tra le quali la allora anonima SS 433 e forniva le coordinate celesti e le mappe necessarie per permettere ad altri astronomi di individuare tali oggetti. Sfortunatamente le coordinate elencate per SS 433 erano errate, fatto probabilmente imputabile a un semplice errore di trascrizione. L'oggetto non potè essere osservato nuovamente anche se era stata destata nuova curiosità nell'osservatore.

Quasi contemporaneamente all'effettuazione dell'esplorazione di Case, SS 433 venne scoperta (senza volerlo) in un programma di osservazioni del tutto indipendente dal primo. I radioastronomi dell'Università di Cambridge stavano compilando uno dei numerosi elenchi generali delle radiosorgenti celesti, in questo caso l'elenco venne pubblicato come il Fourth Cambridge Catalogue. La maggior parte delle sorgenti di radioonde comprese nel catalogo risultarono a un dettagliato esame delle fotografie ottiche associate a oggetti extragalattici lontani quali galassie e quasar. L'emissione radio da stelle normali della nostra galassia è semplicemente troppo debole per poter essere rivelata in tali esplorazioni, anche per oggetti vicini al sistema solare. Nel corso dell'esplorazione venne individuata una radiosorgente piuttosto luminosa dai ricercatori di Cambridge in Aquila e venne indicata come 4C 04.66.

Se qualcuno avesse notato in questo punto la coincidenza della posizione della radiosorgente con la posizione di una stella visibile di debole luminosità, sarebbero state sicuramente intraprese ulteriori osservazioni ottiche in tale punto, a causa della scarsità di radiosorgenti stellari rivelabili visivamente. Tuttavia, ancora una volta, le coordinate celesti pubblicate per 4C 04.66 non coincidevano con la effettiva posizione di SS 433. In questo caso il motivo probabile della discrepanza era di natura strumentale; tale particolare regione del cielo è un confuso e aggrovigliato insieme di radiosorgenti in primo piano e sullo sfondo e gli strumenti usati nell'esplorazione di Cambridge non erano in grado di fissare con precisione neppure la posizione di una sorgente relativamente intensa in mezzo a tale confusione. Era andata perduta ancora una volta l'occasione di osservare la particolare natura di SS 433.

In realtà una parte delle cause di confusione ha in questo caso una diretta incidenza nella storia di SS 433. Nel centro di tale regione del cielo si trova un'estesa zona di emissioni radio che copre una superficie doppia delle dimensioni angolari della Luna piena. Le caratteristiche di tali radioemissioni sono tipiche di un oggetto catalogato in precedenza e ormai dimenticato, chiamato W50. (La designazione deriva dalla comparsa della sorgente in un catalogo di caratteristiche di radioemissioni compilato negli anni cinquanta dal radioastronomo Gart Westerhout dell'Università del Maryland.) La struttura radio e lo spettro radio di W50 hanno indotto la maggior parte dei ricercatori a credere che si tratti di resti di una vecchia supernova: i resti diffusi in espansione di un'antica esplosione stellare, simile alla Nebulosa del Granchio. Dal momento che non esistono documenti storici che indicano l'osservazione di una supernova in tale parte del cielo, è stato tuttavia difficile ricavare una prova assoluta di questa affermazione sulla natura di W50. Se si tratta davvero di resti di una supernova, le sue dimensioni e la sua luminosità radio, confrontate con i dati analoghi ottenuti per resti meglio studiati, fanno pensare che l'esplosione sia avvenuta approssimativamente 10 000 anni fa. Pur se una tale età è chiaramente troppo elevata perché l'esplosione possa essere stata registrata da osservatori umani, essa qualifica ugualmente l'evento che ha creato W50 come una delle supernove più recenti.

La mappa radio di una grande area attorno a SS 433 rivela che tale oggetto è immerso in un'estesa sorgente di emissioni radio, chiamata W50, forse i resti di un'antica esplosione di supernova. La mappa, che copre una regione del cielo di circa due gradi, è in codice di colore in modo tale che il rosso rappresenta la emissione radio più intensa e il blu la meno intensa. La macchia luminosa rossa presso il centro coincide con SS 433. La analoga macchia verso nord-est *(in alto a sinistra)* è probabilmente una radiosorgente extragalattica compatta, vista attraverso il bordo di W50. La radiosorgente estesa verso nord-ovest *(in alto a destra)* è una nube isolata di idrogeno ionizzato tra W50 e il sistema solare. I dati furono registrati a una lunghezza d'onda di 11 centimetri da B. J. Geldzahler, T. Pauls e C. J. Salter, che lavoravano con il radiotelescopio da 100 metri dell'Istituto di radioastronomia Max Planck di Effelsberg nella Germania Occidentale.

Una mappa radio ad alta risoluzione dell'area nelle immediate vicinanze di SS 433 è stata realizzata a una lunghezza d'onda di sei centimetri con il Very Large Array, un complesso di radiotelescopi presso Socorro, nel New Mexico. Le due strutture allungate che partono dalla sorgente puntiforme centrale della mappa sono allineate con le protuberanze della radiosorgente circostante, W50. La mappa è stata realizzata in collaborazione da John T. Stocke dell'Università dell'Arizona e da Ernest R. Seaquist e William S. Gilmore dell'Università di Toronto. La mappa a colori è stata approntata da Eric W. Greisen del National Radio Astronomy Observatory. L'ellitticità della sorgente centrale è una conseguenza del processo di rappresentazione delle immagini.

L'intensa sorgente di raggi X coincidente con la posizione di SS 433 appare al centro di questa immagine prodotta da un elaboratore costruita con i dati ottenuti da un telescopio a raggi X a bordo del satellite *HEAO-2*, chiamato anche Osservatorio Einstein. La debole evidenza di raggi X emessi verso l'esterno da entrambe le parti della sorgente centrale fa pensare alla presenza di getti di gas caldo in espansione verso le estremità est e ovest di W50. Questa immagine a raggi X, risultato dell'esposizione di cinque ore, è stata realizzata da Seaquist e Gilmore insieme a Jonathan E. Grindlay e Frederick D. Seward del Centro di astrofisica dell'Harvard College Observatory e dello Smithsonian Astrophysical Observatory. Il campo visivo è di circa un grado.

Una serie di osservazioni radio progettate per costruire una mappa della struttura dettagliata di W50 è stata pubblicata da David H. Clark e dai suoi collaboratori all'Università di Sydney. La mappa riporta anche la sorgente radio luminosa e puntiforme oggi nota come SS 433, circondata dalla diffusa emissione radio di W50. Oggi, a impedire ulteriori studi di questa curiosa configurazione, è intervenuto un altro fatto casuale. Anche se la effettiva struttura radio di W50 è del tutto simmetrica, con SS 433 collocata in prossimità del centro la mappa pubblicata da Clark e dai suoi colleghi mostrava soltanto la parte settentrionale dei resti, nascondendo in questo modo in una certa misura l'importante posizione centrale di questa misteriosa sorgente puntiforme. Se altri osservatori avessero apprezzato la sorprendente simmetria dei resti e avessero quindi riconosciuto che la radiosorgente puntiforme non identificata era centrata con estrema precisione, sarebbero state quasi sicuramente stimolate ulteriori osservazioni.

Anche se si crede universalmente che le supernove lascino spesso (forse sempre) dietro di sé un singolare resto stellare collassato quale una stella di neutroni o un buco nero, vi sono soltanto due casi indiscutibili nei quali tra i resti di una supernova è stata effettivamente scoperta una stella collassata. I resti visibili e radioemittenti sia nell'evento del Granchio che di un altro nella costellazione della Vela comprendono anche un pulsar, il che prova l'esistenza di una stella di neutroni rotante. Il fatto che si conoscano soltanto queste due coincidenze, nonostante l'esistenza di dozzine di resti di supernova emittenti luce e radioonde e di centinaia di radiopulsar, è scoraggiante anche se probabilmente non significativo. L'opportunità di esplorare un'altra coincidenza potenziale di questo tipo, qualora si fosse presentata, non sarebbe stata lasciata sfuggire.

Il preludio finale alle indagini sulle strane proprietà di SS 433 riguarda la sua emissione di raggi X. Come nel caso delle emissioni radio, pochissime stelle normali sono sorgenti rivelabili di raggi X, pur se anche in questo caso si è scoperto che molti resti di supernova sono sorgenti di raggi X. All'inizio degli anni settanta due satelliti terrestri orbitanti hanno registrato indipendentemente emissioni di raggi X nelle vicinanza di Aquila, e ancora una volta la sorgente di queste emissioni, oggi nota come SS 433, fu regolarmente battezzata e catalogata. Gli osservatori che lavoravano con i dati del satellite americano *Uhuru* designarono la sorgente di raggi X con la sigla 4U1908 + 05 (dal quarto catalogo di *Uhuru*, con le coordinate celesti approssimate della sua posizione nel cielo). Nel frattempo un numeroso gruppo britannico (guidato, però, da un americano, Frederick D. Seward), lavorando con i dati del satellite britannico *Ariel V*, chiamarono l'oggetto A1909 + 04.

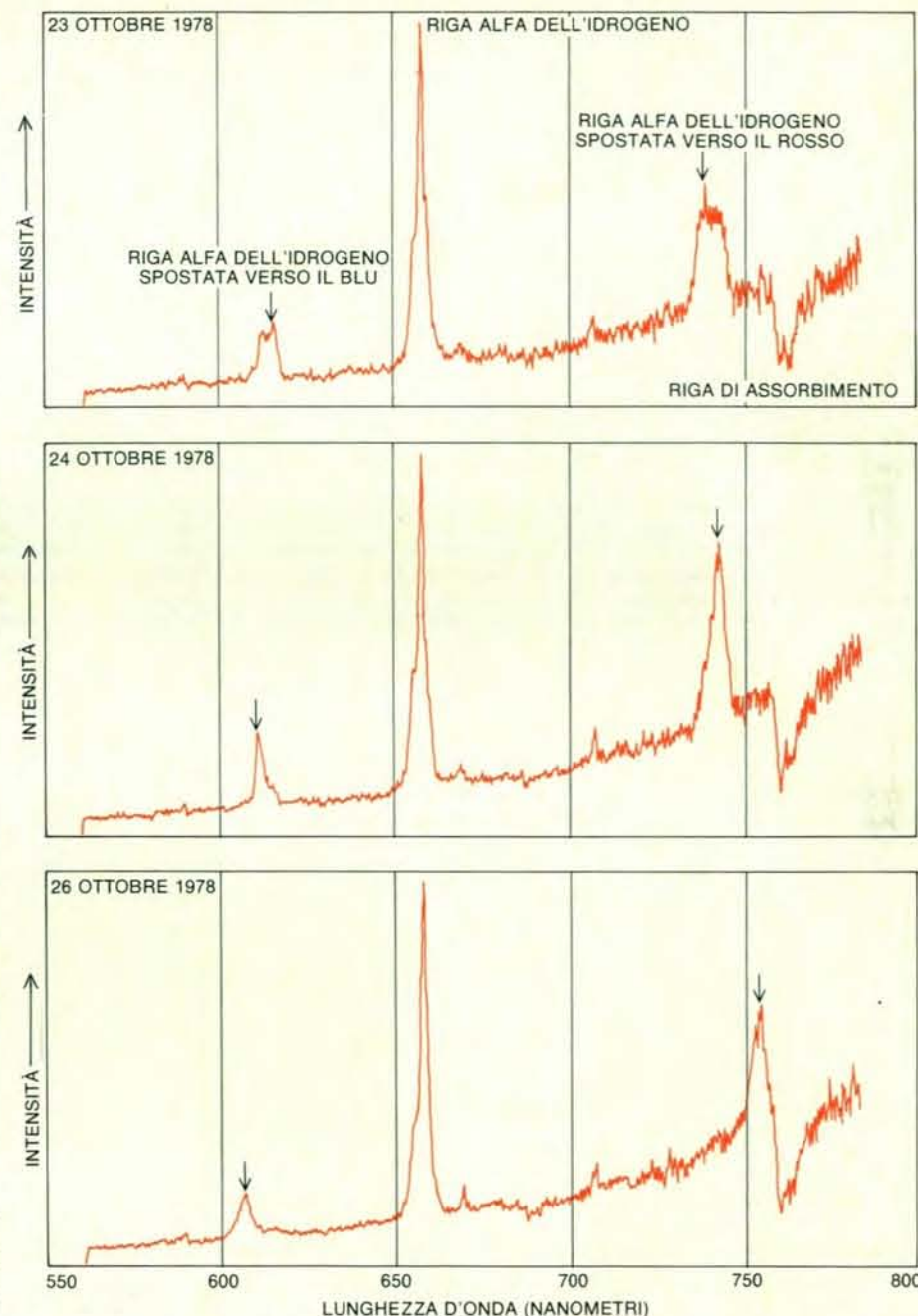
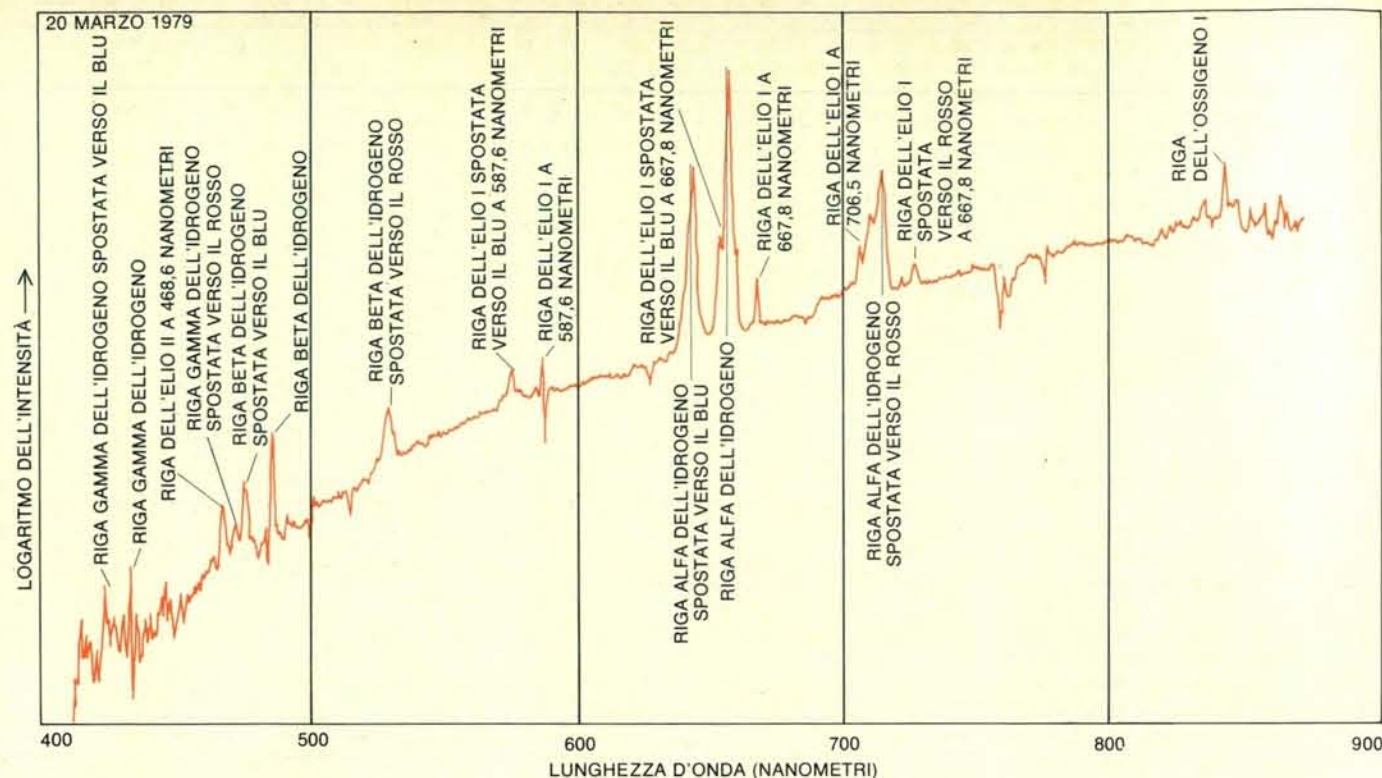
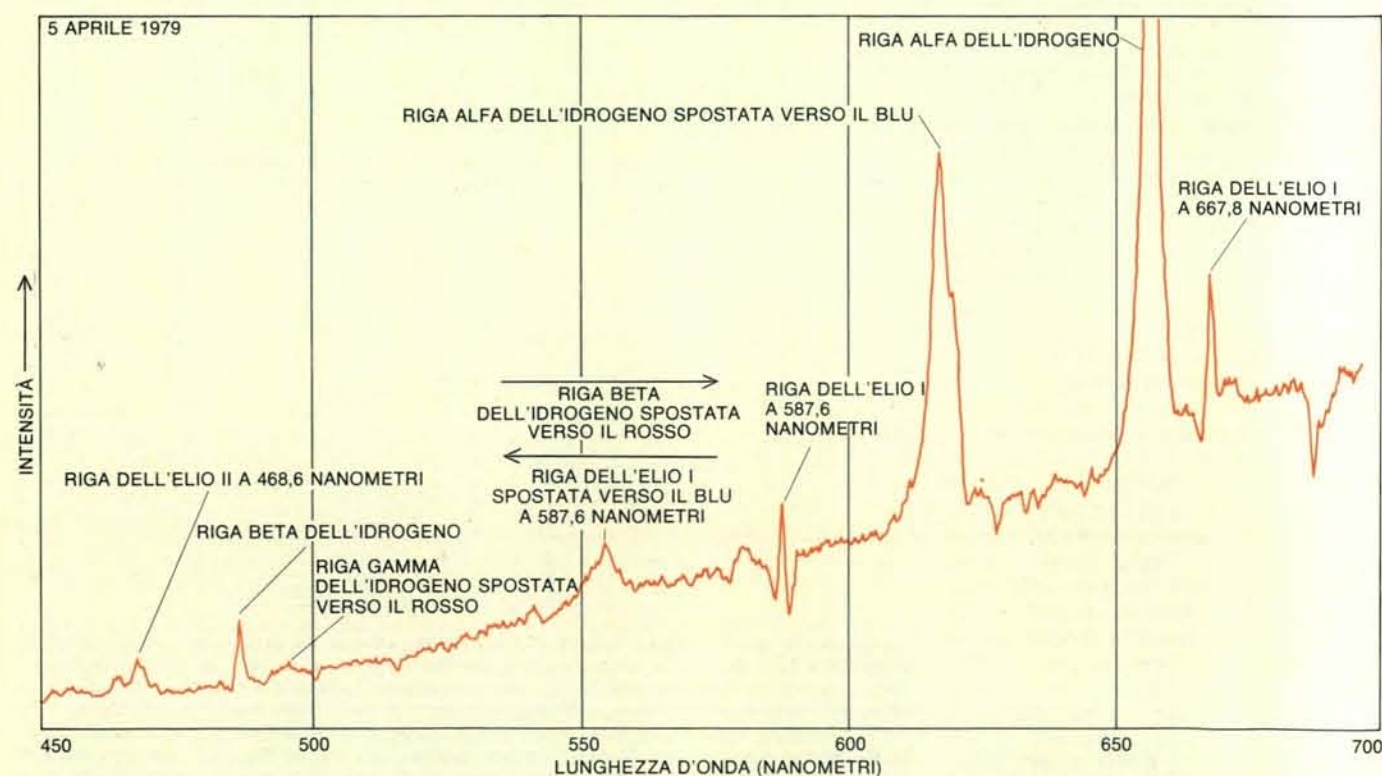
La scarsa risoluzione spaziale degli esperimenti impedì a entrambi i gruppi di accorgersi della coincidenza tra la sorgente di raggi X e la stella luminosa. Seward e i suoi collaboratori fecero però notare che l'intensità dei raggi X sembrava variare nel tempo, caratteristica questa non osservata nell'emissione di raggi X da resti di supernova, a causa della loro relativamente lenta evoluzione dopo l'esplosione iniziale. Seward e i suoi colleghi britannici commentarono profeticamente nella loro pubblicazione del 1976 che probabilmente la sorgente di raggi X non era semplicemente il residuo W50, ma probabilmente qualcosa di più strano a esso collegato.

Una sintesi di queste numerose indicazioni sulle insolite caratteristiche di SS 433 è emersa alla fine nell'estate del 1978, come risultato degli sforzi simultanei, ma del tutto indipendenti, di tre diversi gruppi di ricerca. Un gruppo di radioastronomi canadesi guidati da Ernest R. Seaquist dell'Università di Toronto compì ricerche sulle emissione da giovani stelle eseguendo una nuova esplorazione di oggetti del catalogo Stephenson-Sanduleak. Anche se ciò portò a un'ulteriore riscoperta della radiosorgente, Seaquist e i suoi colleghi si accorsero giustamente che la sorgente era associata al luminoso oggetto stellare SS 433. Nel frattempo un gruppo di radioastronomi britannici e australiani guidato da Sir Martin Ryle di Cambridge affrontava il problema del tutto separato della osservata scarsità di radiosorgenti puntiformi all'interno degli estesi resti di supernova. In una esplorazione condotta con strumenti molto sensibili allo scopo di trovare e controllare le posizioni di tali oggetti, essi scoprirono nuovamente la radiosorgente in W50, e furono in grado a loro volta di osservare la coincidenza con l'oggetto ottico di 14ª grandezza.

Infine, Clark e il suo collega Paul Murdin decisero di ricavare spettrogrammi di stelle visibili la cui posizione era vicina alla radiosorgente piuttosto imprecisamente localizzata che Clark aveva precedentemente notato nella sua mappa di W50. Lavorando in Australia con il telescopio anglo-australiano, essi registrarono lo spettro di SS 433 nel giugno 1978, le prime osservazioni spettroscopiche da quando erano state sviluppate le prime imprecise lastre con il prisma obiettivo nel catalogo di Stephenson-Sanduleak. Clark e Murdin non si accorsero subito che stavano osservando un oggetto già catalogato. Ciò nonostante, lo spettrogramma che essi ricavarono non lasciava dubbi sul fatto che essi avessero opportunamente localizzato la contropparte visibile della strana radiosorgente. Le loro osservazioni mostravano righe di emissione con un'intensità osservata soltanto nelle stelle più insolite. Una localizzazione radio estremamente precisa fornita dal gruppo di Ryle identificò definitivamente SS 433 con la radiosorgente; le posizioni dell'oggetto radio e di quello ottico coincidevano esattamente. Un legame definitivo e indiscutibile con la sorgente X è stato sancito in una recente serie di misurazioni estremamente precise dei raggi X eseguita da Seaquist e dal suo gruppo con il satellite terrestre orbitale *HEAO-2*, chiamato anche Osservatorio Einstein; oggi si sa che la posizione della sorgente X coincide esattamente anche con la posizione della radiosorgente e con quella ottica.

Clark e Murdin pubblicarono le loro osservazioni in una breve nota su «Nature» nell'autunno 1978. Essi identificarono le principali righe di emissione scoprendo in esse le lunghezze d'onda adatte per eccitare atomi di idrogeno e di elio (la quale cosa non è di per sé sorprendente, dal momento che questi elementi sono invariabilmente quelli più abbondanti nelle stelle). Essi accennarono anche indirettamente alla presenza nello spettro di righe di emissione più deboli di origine incerta. (Si vide in seguito che questi aspetti brevemente menzionati costituivano la soluzione completa del mistero di SS 433.) Clark e Murdin, infine, evidenziarono la sorprendente coincidenza tripla di una stella a riga di emissione visibile, di una sorgente di raggi X e di una radiosorgente, tutte situate nel centro dei resti di una supernova e suggerirono che SS 433 e W50 potevano essere causalmente correlate.

A questo punto ebbe inizio il mio personale rapporto con SS 433. Ero da tempo interessato alle caratteristiche ottiche delle deboli contropparti visibili di sorgenti di raggi X celesti. La maggior parte delle stelle a raggi X della nostra galassia appartengono a sistemi binari formati da un oggetto compatto quale una nana bianca o una stella di neutroni legata in un'orbita chiusa a una stella relativamente normale, spesso non troppo diversa dal Sole. L'attrazione gravitazionale esercitata dalla stella compatta sulla sua compagna spesso fa sì che una stella altrimenti normale perda una parte della sua massa trasferendola all'oggetto compatto. Nel processo di trasferimento la materia effluente raggiunge spesso temperature e densità sufficienti a provocare abbondanti emissioni di raggi X. Studi spettroscopici della stella normale in tali sistemi forniscono frequentemente dati importanti sulle caratteristiche di entrambi gli oggetti che ne fanno parte. Le elevatissime densità della materia in oggetti compatti quali le stelle di neutroni non si possono ottenere in laboratori terrestri e di conseguenza i dati astrofisici forniscono virtualmente le sole informazioni dirette sul comportamento della materia in questi singolari stati.

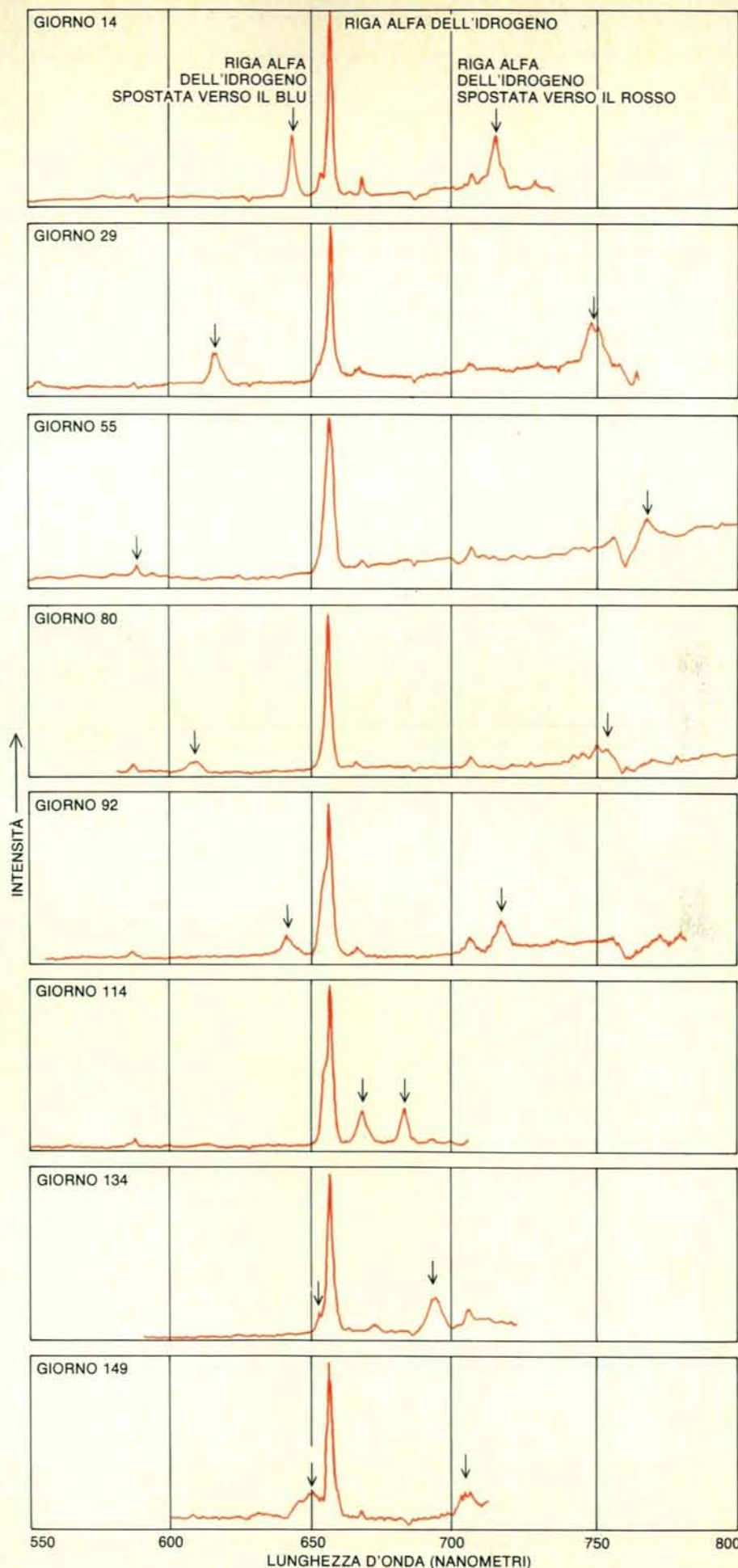
Ricavai il mio primo spettrogramma di SS 433 nel settembre 1978, lavorando con il telescopio Shane di 3 metri al Lick Observatory dell'Università della California. Il mio primo intento era semplicemente quello di confermare i risultati di Clark e Murdin. I dati registrati elettronicamente rivelarono le intense righe di emissione dell'idrogeno e dell'elio osservate da tali ricercatori e anche da Stephenson e Sanduleak. Tuttavia, con mia sorpresa, comparivano nello spettro anche marcate righe di emissione che non mi erano familiari. Era un fatto inquietante. Le righe spettrali che appaiono in uno spettro stellare dipendono dall'abbondanza, dalla temperatura e dalla densità dei singoli elementi che compongono la

Lo spettro di SS 433 è stato registrato in tre diverse notti in un periodo di quattro giorni da Remington P. S. Stone, che lavorava con il telescopio riflettore da 24 pollici dell'Osservatorio Lick. L'aspetto più rimarchevole, il picco a una lunghezza d'onda di 656,3 nanometri, corrisponde all'intensissima riga di emissione dell'idrogeno alfa. Appaiono rispettivamente a 587,6, 667,8 e 706,5 nanometri righe di emissione molto più deboli attribuibili all'elio. Due righe di emissione molto intense, oggi conosciute come versioni spostate per effetto Doppler della riga centrale dell'idrogeno alfa, appaiono a fianco della riga centrale, una *(a sinistra)* verso l'estremità dello spettro a lunghezze d'onda più corte (blu), l'altra verso l'estremità *(a destra)* a lunghezze d'onda più lunghe (rosso). Nel corso delle tre notti la riga spostata verso il rosso ovviamente si allontanava ulteriormente verso il rosso mentre quella verso il blu si avvicinava verso il blu. Il netto minimo nelle curve nei pressi di una lunghezza d'onda di 760 nanometri non ha nulla a che vedere con SS 433; è una riga di assorbimento prodotta dalle molecole dell'atmosfera terrestre.

I tripletti di righe di emissione in questo particolare spettrogramma di SS 433 hanno convinto l'autore e i suoi colleghi dell'Università della California a Los Angeles che le righe di emissione «mobili» osservate nei vari spettri di questo oggetto devono essere attribuiti all'effetto Doppler. Lo spettrogramma è stato ricavato da Steven A. Grandi dell'UCLA nella notte del 20 marzo 1979, con il riflettore Shane da 3 metri del Lick Observatory. Un certo numero di righe di emissione che sono presenti nello spettro dell'idrogeno e dell'elio sono rappresentate qui tre volte: con un'intensa componente centrale e due componenti laterali spostate per effetto Doppler, una delle quali, a sua volta, è spostata verso il rosso e l'altra verso il blu. I forti assorbimenti che appaiono nello spettro sono ancora una volta non correlati con l'oggetto stellare.

Coincidenze poco chiare appaiono talvolta nel gran numero di righe di emissione mobili spostate per effetto Doppler scoperte negli spettri di SS 433. Per esempio, in questo caso sono state identificate numerose intense righe di emissione mobili e altre stazionarie, tra le quali il picco relativamente largo al centro, formato anch'esso da una riga dell'elio spostata verso il blu in moto verso sinistra e da una riga dell'idrogeno spostata verso il rosso in moto verso destra. Nella notte in cui fu ricavato tale spettro (5 aprile 1979) le due righe si incontrarono muovendosi ciascuna in direzione opposta all'altra. Lo spettro è stato registrato al Lick Observatory da Lawrence H. Aller e Charles D. Keyes dell'UCLA.

stella. Nelle situazioni astrofisiche c'è una certa varietà in tali parametri, ma non è infinita; di conseguenza lo spettroscopista stellare si è abituato alla comparsa di certe righe spettrali familiari. L'incontro di una miscellanea di righe di emissione spettrale a lunghezze d'onda diverse è un'esperienza simile a quella di un guidatore che scopre improvvisamente che la strada di accesso al box ha rampe di uscita completamente nuove.

Le intensità delle righe di emissione non identificate nello spettro di SS 433 erano particolarmente interessanti. Queste righe sono solo leggermente meno intense delle righe familiari dell'idrogeno pure presenti nello spettro. Si potrebbe perciò attribuirle a un elemento chimico con un'abbondanza cosmica confrontabile con quella dell'idrogeno. Dato che si pensa che il 90 per cento di tutti gli atomi nella maggior parte delle stelle siano di idrogeno, non ci può essere un elemento altrettanto abbondante. Potrebbe darsi che anche le righe non identificate siano dovute all'idrogeno e che per un qualsiasi motivo esse siano spostate rispetto alle normali lunghezze d'onda osservate in tutte le altre situazioni di laboratorio e di astrofisica?

Esiste in effetti un meccanismo di spostamento ben noto allo spettroscopista, cioè l'effetto Doppler. È noto che il moto relativo della sorgente rispetto all'osservatore sposta leggermente la lunghezza d'onda percepita in qualsiasi fenomeno ondoso, quale il suono o la luce. Inoltre, l'entità dello spostamento fornisce correttamente la velocità del moto e il senso dello spostamento (verso lunghezze d'onda maggiori o minori) indica se l'oggetto si sta allontanando o avvicinando. La grande maggioranza delle stelle presenta un moto casuale di qualche decina di chilometri al secondo rispetto al Sole, causando uno spostamento della riga spettrale di circa lo 0,01 per cento dai valori normali. Per alcune stelle binarie il moto periodico orbitale può essere 10 volte maggiore di quello, provocando così uno spostamento Doppler proporzionalmente maggiore.

Le righe identificate di SS 433 si trovano però a lunghezze d'onda niente affatto prossime ad alcuna riga dell'idrogeno; se fossero perciò state righe dell'idrogeno soggette a spostamento Doppler, le velocità implicate avrebbero dovuto essere prodigiose. Per la più evidente riga non identificata di questi spettrogrammi, situata nella regione rossa dello spettrovicina a una lunghezza d'onda di 740 nanometri, la velocità necessaria per dare origine allo spostamen-

Una scelta di spettri di SS 433, ricavati dall'autore e dai suoi colleghi in un periodo di 164 giorni, copre l'intero ciclo del moto delle righe di emissione spostate per effetto Doppler. I giorni in cui sono state eseguite le osservazioni sono riportati in relazione a un modello del ciclo di 164 giorni nel grafico di pagina 26.

to dalla riga dell'idrogeno più vicina (a 656,3 nanometri) è di circa 40 000 chilometri al secondo, ovvero più del 10 per cento della velocità della luce! Dal momento che la velocità di fuga dalla Galassia è solamente di poche centinaia di chilometri al secondo, velocità stellari maggiori di questa non sono mai state osservate; uno qualsiasi di tali oggetti abbandonerebbe immediatamente (in una scala di tempi astronomica, evidentemente) l'intera Galassia. In breve, l'ipotesi di una emissione dell'idrogeno con spostamento Doppler sembrava una spiegazione insufficiente delle osservazioni.

Una caratteristica ancor più curiosa dello spettro di SS 433 risultò evidente dopo parecchie notti di ripetuto studio dell'oggetto. Le righe non identificate a lunghezze d'onda non familiari apparvero mutare lunghezza d'onda, e in entità notevole. Per esempio, in un periodo di quattro notti l'intensa riga nella regione rossa dello spettro aumentò la sua lunghezza d'onda di più dell'1 per cento. Se ne fosse stato responsabile l'effetto Doppler, tale discrepanza apparentemente piccola si sarebbe dovuta spiegare con una variazione di velocità di quasi 5000 chilometri al secondo in quei quattro giorni. Inoltre, le variazioni di lunghezza d'onda delle righe non identificate non erano neppure coerenti: alcune delle righe si spostavano verso le lunghezze d'onda maggiori, mentre altre si spostavano contemporaneamente verso lunghezze d'onda minori.

A questo punto divenne chiaro che il comportamento spettrale di SS 433 era notevolmente più strano di quanto previsto dalle precedenti osservazioni. I miei colleghi e io iniziammo perciò all'Università della California a Los Angeles un programma per ricavare almeno una breve osservazione spettrale di SS 433 in qualsiasi notte. Nonostante le rigorose limitazioni del tempo di osservazione con grandi telescopi di ricerca, ciò era attuabile per la relativa luminosità dell'oggetto. Per esempio, la sensibile strumentazione spettroscopica controllata da un elaboratore del riflettore Lick da 3 metri deve servire per l'osservazione di galassie e quasar lontanissimi, oggetti centinaia di volte più deboli di SS 433. Uno spettrogramma di buona qualità di questa stella si può ricavare in circa 10 minuti con telescopi di questa categoria e si possono di conseguenza eseguire frequentemente spettrogrammi senza interrompere programmi di osservazione precedentemente progettati.

Alle osservazioni parteciparono ricercatori provenienti da tutte le quattro sedi dell'Università della California nelle quali sono in corso ricerche di astronomia ottica : Los Angeles, Berkeley, Santa Cruz e San Diego. Più di una dozzina di astronomi, i cui interessi di ricerca principali andavano dalle lontane galassie alle vicine stelle normali, rinunciarono generosamente ai loro già modesti tempi di osservazione per aiutare a controllare l'oggetto. Degni di una particolare menzione sono tre ricercatori che hanno ottenuto lo stesso numero di spettrogrammi che sono riuscito a ottenere io: Steven A. Grandi e Holland C. Ford dell'UCLA e Remington P. S. Stone del gruppo del Lick.

Il nostro gruppo aveva poco tempo da perdere, perché dal dicembre al febbraio di ogni anno la linea di vista di SS 433 è troppo vicina al Sole per osservazioni notturne. Entro la fine del periodo di osservazione del 1978 avevamo notato che le righe spettrali «mobili», come avevamo deciso di chiamarle, attraversavano un ampio intervallo di lunghezze d'onda. Per esempio, la riga di emissione più rossa variava la propria lunghezza d'onda di circa 70 nanometri in 30 giorni, il che, se fosse stato interpretato come uno spostamento Doppler in graduale aumento, avrebbe implicato un aumento costante di velocità da 20 000 a 50 000 chilometri al secondo.

Nel comunicare i nostri dati iniziali alla comunità scientifica abbiamo cercato di mantenerci scettici sull'interpretazione delle righe mobili con lo spostamento Doppler per una varietà di ragioni che sembravano tutte valide in quel momento. Le obiezioni più ovvie furono quelle che ho già ricordato. Tanto le velocità in gioco quanto le variazioni di velocità erano maggiori di un fattore 100 dei corrispondenti valori osservati in qualsiasi altro oggetto stellare. Esistevano, inoltre, difficoltà ancora più sottili per la spiegazione in chiave di spostamento Doppler. Alcune righe mobili si spostavano verso lunghezze d'onda maggiori e altre verso lunghezze d'onda minori, sfidando la semplice interpretazione di una sola nuvola di gas in avvicinamento o in allontanamento.

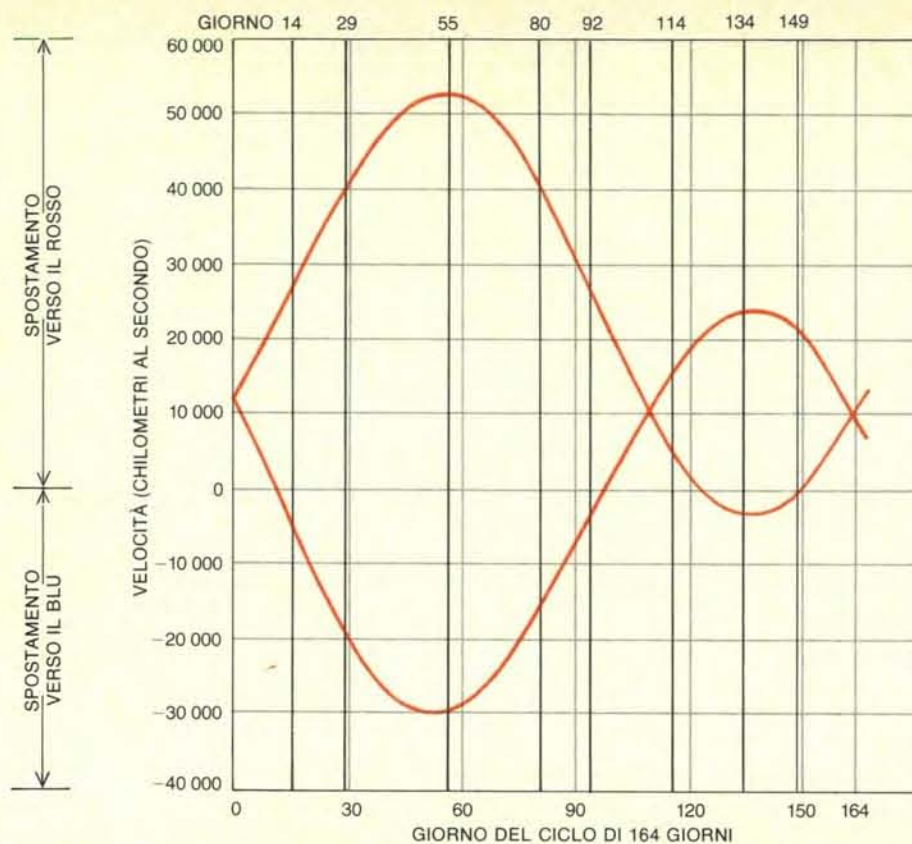
Inoltre, l'identificazione delle righe di emissione mobili con righe di idrogeno sottoposte a spostamento Doppler avrebbe implicato l'esistenza di un gas a una temperatura piuttosto modesta, circa 20 000 kelvin o meno; a temperature superiori la maggior parte dell'idrogeno sarebbe ionizzato e le righe spettrali dovute alle transizioni di un elettrone legato a un nucleo atomico non potrebbero comparire. Tuttavia, qualsiasi meccanismo immaginabile che possa accelerare un gas alle enormi velocità implicate (una notevole frazione della velocità della luce) riscalderebbe il gas a una temperatura ancora maggiore. Un altro modo di presentare questa difficoltà è di far notare che se le energie termica e cinetica dei nuclei di idrogeno nel gas emittente sono all'incirca uguali, come accade spesso in un gran numero di sistemi fisici, le velocità osservate implicano temperature maggiori di 30 miliardi di kelvin, molto maggiori della temperatura dedotta per il gas osservato in SS 433.

In questo grafico sono riassunti due anni di osservazioni, che rappresentano i valori delle righe di emissione spostate verso il rosso e verso il blu osservate nello spettro di SS 433 dalla metà del 1978 alla metà del 1980 in funzione della velocità equivalente del gas espulso. Le grandi interruzioni dei dati corrispondono ai tempi nei quali la linea di vista verso l'oggetto era troppo vicina al Sole per consentire osservazioni notturne; le interruzioni più piccole si possono attribuire alla vicinanza della Luna, che rende difficile l'osservazione, o alla mancanza di tempi di osservazione su un telescopio adatto. Le curve riproducono il comportamento basato sull'ipotesi che il gas venga emesso in due getti in direzioni opposte.

Le nostre conclusioni non scoraggiarono molti teorici dotati di immaginazione. Andrew Fabian e Martin Rees di Cambridge fecero notare che l'esistenza di un gas sottoposto a spostamento Doppler nei due sensi si potrebbe spiegare se ci fosse un oggetto stellare che espelle due getti di gas in direzioni pressoché opposte. In tal caso una certa quantità di gas si avvicinerebbe all'osservatore, mentre l'altra si sta allontanando. Fabian e Rees notarono l'esistenza di siffatte strutture a due lobi osservate su una scala molto più ampia in radiogalassie, nelle quali si trova spesso che l'emissione radio è concentrata in due getti opposti. Uno schema simile è stato proposto indipendentemente da Mordechai Milgrom del Weizmann Institute of Science di Israele, che andò persino oltre avanzando un'ipotesi che ebbe in seguito uno spettacolare successo. Soltanto basandosi su pochi punti dei nostri dati, Milgrom avanzò l'ipotesi che i moti delle righe potessero essere periodici, con un intervallo di ripetizione all'incirca di qualche mese.

Nel frattempo noi continuavamo a preoccuparci di tutta l'idea fondamentale. Le righe di emissione mobili non identificate erano davvero righe di idrogeno sottoposte a spostamento Doppler? La risposta fu chiara nel marzo dello scorso anno, quando SS 433 era nuovamente sufficientemente lontana dal Sole per osservazioni spettroscopiche. Il primo spettrogramma della nuova stagione di osservazione mostrava chiaramente che tutte le principali righe di emissione erano tripletti: una componente alla lunghezza d'onda del laboratorio (non spostato), una a una lunghezza d'onda spostata verso l'estremità rossa dello spettro (cioè verso le lunghezze d'onda maggiori) e una spostato verso l'estremità blu (verso le lunghezze d'onda minori). Questa particolare struttura a tripletti era mostrata dalle righe di emissione sia dell'idrogeno sia dell'elio, in tutto circa una dozzina di righe diverse. Inoltre, tutte le righe spostate verso il rosso implicavano indipendentemente la stessa velocità di recessione, di circa 27 000 chilometri al secondo, mentre tutte le righe spostate verso il blu implicavano allo stesso modo la stessa velocità di avvicinamento, di circa 6000 chilometri al secondo. Questo moltiplicarsi di coincidenze poteva essere spiegato solamente con lo spostamento Doppler. Alla stessa conclusione giunsero indipendentemente e contemporaneamente un gruppo di astronomi dell'Università dell'Arizona diretti da James W. Liebert.

Il fatto curioso è che la difficoltà nell'interpretazione delle righe mobili di SS 433 fu una allucinante ripetizione dell'improvviso riconoscimento del significato degli spettri dei quasar fatto da Maarten Schmidt vent'anni fa. Anche gli enormi spostamenti verso il rosso dei quasar, causati dall'espansione dell'universo, spostavano le righe di emissione dalle loro lunghezze d'onda normali tanto da rendere difficile l'identificazione degli atomi che le originavano, nonostante il fatto che le righe risultavano associate agli elementi più abbondanti e familiari dell'universo. Anche se la storia della scoperta dei quasar era ben nota a tutti noi, SS 433 presentava un problema di interpretazione iniziale attribuibile a due complicazioni senza precedenti: in primo luogo, tali enormi spostamenti verso il rosso non erano mai stati associati a stelle interne alla nostra galassia e in secondo luogo l'oggetto mostrava chiaramente sia spostamenti verso il rosso così come spostamenti verso il blu.

È chiaro che dal momento che la lunghezza d'onda delle righe sottoposte a spostamento Doppler cambia rapidamente nel tempo, sta cambiando sia la velocità del gas che si avvicina sia quella del gas che si allontana. Una domanda ovvia è allora questa: esiste un qualche schema regolare per queste variazioni? Nei nostri due anni di osservazioni abbiamo accumulato circa 150 misurazioni separate sia di righe di emissione spostate verso il rosso sia di righe spostate verso il blu; un grafico dei valori delle velocità in gioco in funzione del tempo rivela numerosi aspetti affascinanti nello strano comportamento delle righe mobili *(si veda l'illustrazione in queste due pagine)*. Si vede subito che le velocità sono realmente enormi: il gas sottoposto a spostamento verso il rosso (in allontanamento) raggiunge valori fino a 50 000 chilometri al secondo (il 16 per cento della velocità della luce) in molte occasioni differenti, mentre il gas sottoposto a spostamento verso il blu (in avvicinamento) raggiunge velocità fino a 30 000 chilometri al secondo. Poiché le velocità delle stelle della nostra galassia (sia in avvicinamento sia in allontanamento) non superano mai qualche centinaio di chilometri al secondo e poiché tutti gli oggetti extragalattici appena al di là della nostra galassia mostrano soltanto spostamenti verso il rosso, attribuibili all'espansione dell'universo, SS 433 è contraddistinta dal mostrare il massimo spostamento verso il blu (di un fattore 100) riscontrabile tra tutti gli

Le curve teoriche tracciano un ciclo completo di 164 giorni nello schema previsto di righe di emissione di SS 433 spostate verso il rosso e verso il blu, nell'ipotesi che il gas che emette luce sia concentrato in due getti rotanti diretti in versi opposti, ognuno dei quali con una velocità di espulsione di 78 000 chilometri al secondo, pari approssimativamente a un quarto della velocità della luce. Le righe nere indicano i giorni nei quali sono stati ottenuti gli spettri selezionati che sono stati riportati nel grafico di pagina 23. Le caratteristiche di emissione degli spettri coincidono con notevole precisione con gli spostamenti verso il rosso e verso il blu previsti. Le curve mostrano uno spostamento medio verso il rosso costante equivalente a una velocità di espulsione di 12 000 chilometri al secondo, dovuto all'effetto della relatività ristretta chiamato dilatazione dei tempi.

oggetti celesti conosciuti, galattici o extragalattici.

Altri pezzi dell'enigma andarono a posto dopo un'analisi del modo in cui cambiavano le velocità. I volumi in avvicinamento e in recessione del gas radiante non coincidono sicuramente nello spazio; se così fosse, essi si separerebbero rapidamente a tali enormi velocità dirette in senso opposto. Tuttavia, nonostante la loro separazione fisica le due nubi di gas sono correlate in modo ben definito; le variazioni nei sistemi sottoposti a spostamento verso il rosso e verso il blu raggiungono le loro velocità massime contemporaneamente. Inoltre, il valore medio delle due velocità in una notte qualsiasi è pressoché costante, ma molto elevato, pari a circa 12 000 chilometri al secondo.

La costanza del valore medio delle due velocità per effetto Doppler, nonostante l'enorme variazione dei loro valori singoli in una scala di tempi di giorni, si potrebbe forse capire se un oggetto centrale fosse responsabile dell'espulsione di entrambe le nubi radianti; in tal caso la velocità media sarebbe quella della stella centrale. Tuttavia, ancora una volta, ci trovammo di fronte al problema che tale valore supera con ampio margine la velocità di fuga dalla nostra galassia; l'oggetto si allontanerebbe indefinitamente dalla nostra galassia in una piccolissima frazione dell'età delle stelle nelle vicinanze del Sole (almeno 10 miliardi di anni). È ragionevole supporre di avere la fortuna di vivere e di essere in grado di eseguire osservazioni astronomiche proprio durante questo così breve intervallo? Risulta disponibile una spiegazione decisamente meno forzata.

Devo ancora precisare quella che è forse la più sorprendente caratteristica della variazione delle velocità del materiale associato a SS 433. Nonostante le grandi e continue variazioni di velocità e la notevole distanza di tempo tra le osservazioni, si può vedere, dai grafici delle due pagine precedenti che tanto il gas sottoposto a spostamento verso il rosso quanto quello sottoposto a spostamento verso il blu sono periodicamente tornati agli stessi valori di velocità all'incirca ogni sei mesi; in altre parole, le variazioni di velocità sono periodiche. Una semplice analisi dei dati attuali mostra che il periodo esatto è di 164 giorni, con una indeterminazione di circa mezza giornata. Lo schema delle variazioni è chiaramente caratterizzante. Due volte ogni 164 giorni i due sistemi emittenti raggiungono la stessa velocità e poi si scambiano ruolo, ovvero il gas sottoposto a spostamento verso il blu si sposta verso il rosso e viceversa. Dal punto di vista osservativo tali eventi appaiono nello spettro come una comparsa graduale delle righe mobili che poi si incrociano facendosi da parte in direzioni opposte. Abbiamo osservato più volte questo evento di «attraversamento», anche se non siamo stati ancora sufficientemente fortunati da ricavare uno spettrogramma durante l'esatto (e presumibilmente breve) intervallo durante il quale le righe mobili si sovrappongono esattamente.

Esiste un ben noto precedente a questo genere di variabilità spettrale periodica, anche se unico su una scala fortemente ridotta. In certi sistemi di stelle binarie, se la luminosità delle due stelle è paragonabile, si vedono talvolta due gruppi di righe spettrali indipendenti. Dato che in ogni istante una delle stelle ha una certa componente di velocità verso il sistema solare, mentre l'altra stella ha una certa componente in direzione opposta, gli spostamenti Doppler della stessa riga spettrale in ciascuna stella sono leggermente differenti; quindi le righe appaiono come una coppia ben risolta a lunghezze d'onda lievemente differenti. Quando le due stelle ruotano attorno al loro comune centro di massa la componente del moto di ciascuna stella verso o lontano dalla Terra varia in modo regolare e periodico; di conseguenza le lunghezze d'onda di ognuna delle due righe spettrali mostrano una variazione periodica tra un limite a bassa lunghezza d'onda e uno a elevata lunghezza d'onda. Nei sistemi di stelle binarie, però, l'ampiezza delle variazioni di velocità non è mai maggiore di qualche centinaio di chilometri al secondo, mentre per SS 433 l'ampiezza è più di 100 volte maggiore.

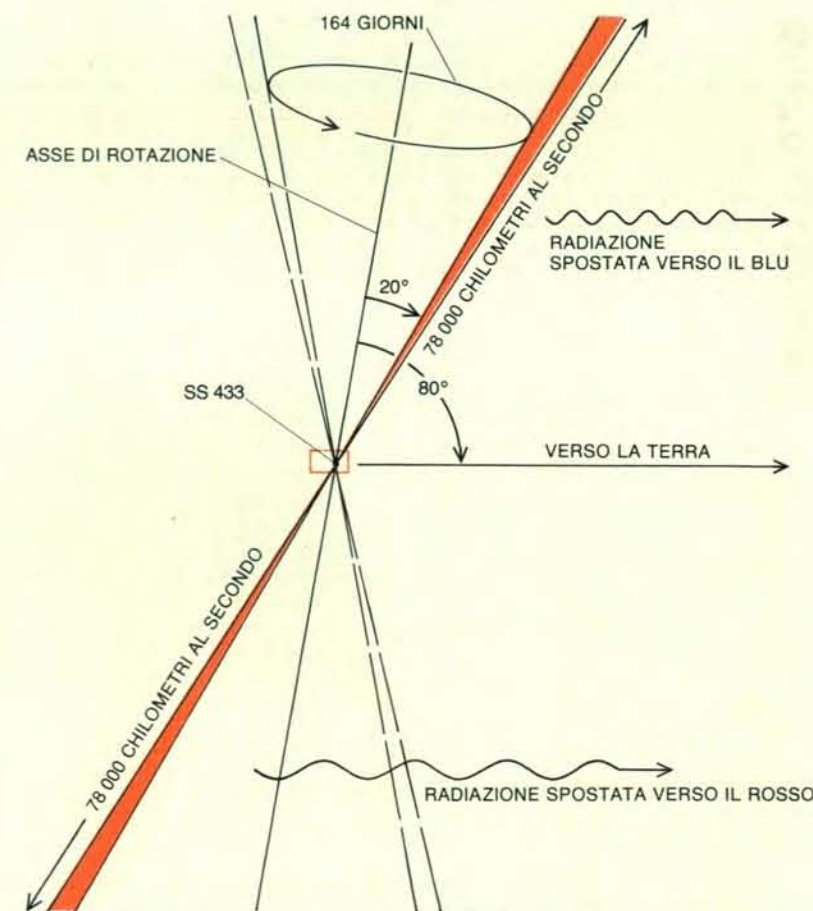
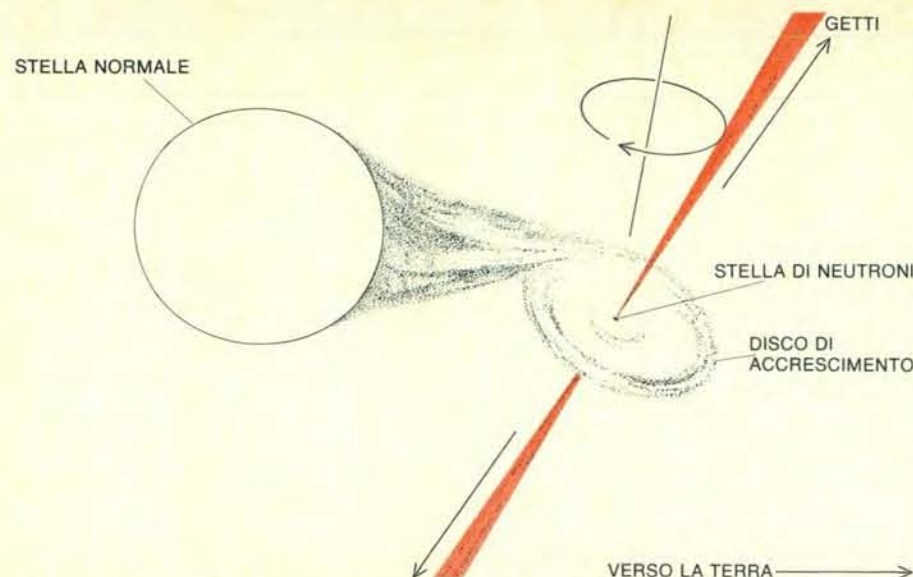
Potrebbe SS 433 rappresentare qualche versione enormemente amplificata di questo fenomeno, formata semplicemente da due oggetti orbitanti uno rispetto all'altro? Se le cose stessero così, vi sarebbe una conseguenza interessante. Un corollario della semplicità della forza gravitazionale è che tutti i corpi orbitanti, indipendentemente dalla loro natura, seguono una relazione fondamentale tra il periodo orbitale, la velocità orbitale e la massa totale del sistema. Se si suppone perciò che le drastiche variazioni di lunghezza d'onda di SS 433 siano dovute al moto orbitale, allora si può direttamente calcolare la quantità totale di materia in SS 433 soltanto dall'osservazione dell'ampiezza e del periodo della variazione di velocità. Il risultato è una massa equivalente a un miliardo di volte la massa del Sole! Dal momento che le stelle conosciute più pesanti hanno una massa inferiore a 100 masse solari, questo risultato è stupefacente. Infatti, la massa attribuita in questo modo a SS 433 è circa l'1 per cento della massa totale dei 100 miliardi di stelle di tutta la Galassia.

Si può pensare che un oggetto con una massa così elevata possa esser stato finora trascurato? La risposta è no. Esistono diverse argomentazioni differenti che si possono tirare in ballo per dimostrare che le strane variazioni delle righe spettrali non possono essere dovute al moto orbitale. Per esempio, i parametri orbitali stimati sono tali che il diametro dell'orbita dovrebbe essere talmente grande che la luce impiegherebbe circa due settimane per andare dall'uno all'altro degli oggetti orbitanti. Tuttavia le creste e i ventri del grafico degli spostamenti verso il rosso e verso il blu del sistema rimangono sincronizzati uno con l'altro con una precisione di circa un giorno. Secondo la teoria della relatività ristretta, nessuna informazione, indipendentemente dal metodo di trasmissione, può propagarsi con velocità maggiore di quella della luce. Risulterebbe perciò impossibile per i due oggetti rimanere sincronizzati; in altri termini, nessuno dei due può «dire» all'altro dove si trova.

Un secondo argomento che parla chiaramente a sfavore dell'ipotesi del moto orbitale è il sistema di emissione stazionaria dello spettro di SS 433, cioè il gruppo di righe spettrali che sono costantemente vicine alle loro lunghezze d'onda di laboratorio. Se il gas che emette tali righe si trovasse in prossimità di un oggetto con una massa pari a un miliardo di masse solari, sentirebbe subito tale enorme attrazione gravitazionale e cadrebbe verso la massa maggiore. Invece le velocità indicate da queste righe sono decisamente piccole.

Per tutti questi motivi, uniti ad argomentazioni più raffinate, non è credibile l'ipotesi di due corpi orbitanti uno rispetto all'altro per spiegare le rapide variazioni di lunghezza d'onda dello spettro di SS 433. La maggior parte dei ricercatori interessati al problema si sono invece orientati verso una modifica del concetto secondo il quale un oggetto centrale espelle due getti di materia, uno diretto all'incirca verso la Terra, quindi originante una radiazione con riga spostata verso il blu, e l'altro diretto approssimativamente lontano dalla Terra e quindi originante la radiazione spostata verso il rosso. Se poi si postula che la retta immaginaria congiungente i due getti ruoti a una velocità tale da compiere un giro completo ogni 164 giorni, si spiegano anche le modulazioni periodiche dei valori di velocità osservate. Ciò è dovuto al fatto che in ogni punto del ciclo di 164 giorni l'angolo di ogni getto rispetto alla linea di vista della Terra varia. Quando i getti sono vicini a essere orientati direttamente verso la Terra (o lontano da essa), si osservano le massime velocità di avvicinamento (o di recessione). Quando invece i getti sono orientati perpendicolarmente alla linea di vista, non c'è alcun gas che si muove verso la Terra o lontano da essa e ci si aspetta che le velocità abbiano il valore minimo.

È facile quantificare tale ipotesi con un semplice insieme di equazioni. Sorprendentemente vi sono soltanto cinque incognite nelle equazioni; in altre parole, dovrebbero bastare cinque parametri caratteristici per descrivere completamente il comportamento dell'oggetto. Tanto per cominciare vi sono due angoli geometrici incogniti: uno è l'inclinazione dell'asse di rotazione dell'oggetto rispetto alla linea di vista. Per esempio, un osservatore extraterrestre capace di accorgersi della rotazione terrestre non è detto che debba necessariamente essere in quiete al di sopra dell'equatore; egli è in grado di vedere la rotazione da qualsiasi latitudine. Noi non conosciamo la «latitudine» dalla quale stiamo osservando SS 433. Il secondo angolo incognito è l'inclinazione dell'asse dei getti rispetto all'asse di rotazione. Non è detto che il nostro angolo rispetto ai getti debba coincidere con il primo angolo, allo stesso modo in cui gli assi dei poli terrestri geografici e magnetici sono lievemente inclinati uno rispetto all'altro. Una terza incognita è la velocità di espulsione dei getti; per semplicità conviene supporre che i due getti siano orientati in direzioni opposte, ma con la stessa velocità. (Se tale ipotesi è errata, non vi sono soluzioni delle equazioni.) Infine i valori esatti del periodo e della fase del ciclo di 164 giorni di SS 433 sono incognite che vanno determinate a partire dai dati.

Le nostre osservazioni spettroscopiche si possono adattare a questo modello teorico per stabilire se esistono valori dei cinque parametri che sembrano adattarsi ai dati. Risulta, forse sorprendentemente, che questa semplice ipotesi si adatta benissimo ai dati. I due angoli sono risultati rispettivamente di circa 80 e 20 gradi e la velocità dei getti secondo le osservazioni è di 78 000 chilometri al secondo, ovvero il 26 per cento della velocità della luce. Il motivo per cui non abbiamo mai osservato direttamente uno spostamento verso il rosso o verso il blu di tale entità è semplicemente che i due angoli geometrici non risultano essere angoli retti. Quindi i getti non sono orientati direttamente verso o lontano dalla Terra. (In realtà, ci sarebbe da sospettare se lo fossero.) La massima componente di velocità verso la Terra o lontano da essa non può perciò raggiungere la effettiva velocità ipotizzata per i getti. Un osservatore situato per caso lun-

La geometria del modello a getto rotante di SS 433 si basa su questo diagramma schematico. L'asse di rotazione dell'oggetto è inclinato sulla linea di vista del sistema solare di un angolo di circa 80 gradi. I getti stessi sono a loro volta inclinati di circa 20 gradi rispetto all'asse di rotazione. Il periodo di rotazione è di 164 giorni. In questo modello la velocità dei getti risulta essere di circa 78 000 chilometri al secondo. La componente di tale velocità proiettata lungo la linea di vista varia periodicamente mentre l'asse dei getti ruota; quindi le righe di emissione dello spettro sottoposte a spostamento Doppler fluttuano tra posizioni massime e minime, a seconda dell'inclinazione dei getti sulla linea di vista. Per esempio, nella posizione qui indicata sia gli spostamenti verso il rosso sia quelli verso il blu dovrebbero trovarsi in un massimo; si dovrebbero osservare spostamenti minimi sia verso il rosso sia verso il blu quando i getti sono perpendicolari alla linea di vista.

Gli ipotetici oggetti centrali postulati dall'autore e dai suoi colleghi per aiutare a spiegare le caratteristiche osservate nello spettro di SS 433 secondo il modello a getto rotante sono illustrati in questo diagramma, che può essere visto come un ingrandimento dell'area all'interno della piccola scatola colorata della figura della pagina precedente. Si pensa che l'oggetto responsabile dell'emissione dei getti faccia parte di un sistema di stelle binario, formato da una stella relativamente normale *(a sinistra)*, non dissimile dal Sole, legata in un'orbita chiusa a una stella di neutroni compatta *(a destra)*, che compare nel processo di espulsione di materiale dalla sua normale compagna in virtù del suo intenso campo gravitazionale. (Si suppone che entrambe le stelle abbiano la stessa massa del Sole.) Il gas che fuoriesce dalla stella normale forma un disco rotante in continua crescita attorno alla stella di neutroni ed è dalle facce di questo disco che vengono espulsi i due getti in direzioni opposte. La precessione del piano del disco di accrescimento che avviene attorno alla stella di neutroni è probabilmente la causa della rotazione dell'asse dei getti.

go la proiezione della congiungente i getti vedrebbe sia la velocità massima sia quella minima.

Per la materia in moto con una significativa frazione della velocità della luce, gli effetti previsti dalla teoria della relatività ristretta diventano importanti. È infatti proprio un effetto del genere che spiega uno degli aspetti più enigmatici di SS 433: la grandissima e costante velocità media (12 000 chilometri al secondo) dei due getti. Tale notevole effetto ha diversi nomi tecnici, quale spostamento Doppler del secondo ordine o trasversale, ma è spesso citato semplicemente come dilatazione dei tempi. Un osservatore ipotetico che stia guardando una persona in moto ad alta velocità fornita di orologio avrà la sensazione che il suo orologio stia rallentando; maggiore è la velocità, più lento è l'orologio. La persona che trasporta l'orologio vedrà invece l'orologio segnare il tempo esatto. In effetti, dato che tutti i moti sono relativi, egli potrebbe attribuire all'osservatore «in quiete» una errata percezione del tempo.

Che cosa ha a che fare tutto ciò con SS 433? Ogni atomo del gas espulso è l'equivalente di un orologio, dal momento che deve registrare la frequenza (o, il che è equivalente, la lunghezza d'onda) per poter emettere luce alla lunghezza d'onda opportuna tutte le volte che i suoi elettroni si diseccitano. Dal momento che la dilatazione dei tempi rallenta sempre l'orologio e dal momento che la frequenza e la lunghezza d'onda della luce sono inversamente proporzionali, la diminuzione di frequenza apparirà come un aumento di lunghezza d'onda, in altri termini come uno spostamento verso il rosso. L'entità dello spostamento verso il rosso dovuto alla dilatazione del tempo dipende soltanto dalla velocità dell'orologio ed è facile calcolare che a una velocità di circa un quarto della velocità della luce l'effetto della dilatazione dei tempi è del 4 per cento.

Ciò può apparire una discrepanza di piccola entità, ma ricordiamo che uno spostamento verso il rosso del 4 per cento della velocità della luce (300 000 chilometri al secondo) è 12 000 chilometri al secondo. Ovviamente, questo è esattamente il valor medio osservato per i due fasci di SS 433. Lo spostamento verso il rosso dovuto alla dilatazione dei tempi è sempre presente nei getti, indipendentemente dal loro angolo rispetto alla linea di vista. Perciò lo spostamento Doppler con variazione periodica di 164 giorni è sovrapposto allo spostamento verso il rosso dovuto alla dilatazione dei tempi (nel senso che si somma o si sottrae da esso). Questa interpretazione spiega elegantemente non solo perché la velocità media effettivamente osservata è di 12 000 chilometri al secondo, ma anche perché le righe di emissione associate ai due getti compaiono due volte ogni 164 giorni a tale grande valore anziché a velocità nulla. Gli eventi di attraversamento si verificano quando i due getti sono orientati perpendicolarmente alla linea di vista. Anche se in tali istanti non c'è né avvicinamento né recessione del gas emittente, e quindi nessuno spostamento verso il rosso o verso il blu che non sia quello dovuto alla dilatazione dei tempi, l'effetto onnipresente della dilatazione dei tempi dà a entrambi i fasci uno spostamento verso il rosso equivalente a 12 000 chilometri al secondo.

Qual è il misterioso oggetto che emette i getti? È difficile dirlo, ma possediamo un'importante indicazione, cioè la velocità osservata: il 26 per cento della velocità della luce. Perché tale valore e non qualche altro? La risposta può essere che tale velocità è molto prossima alla velocità di fuga della materia dalla superficie di una stella di neutroni. I calcoli teorici mostrano che esiste un limitato intervallo di parametri nel quale queste stelle esotiche possono sostenere la loro stessa massa ed essere quindi stabili. Una stella di neutroni con massa uguale a quella del Sole dovrebbe avere un raggio di solo circa 10 chilometri. La velocità necessaria per sfuggire dalla superficie di una stella di neutroni risulta essere simile a quella osservata nei getti di SS 433. Forse si tratta solamente di una coincidenza, ma se non è così essa fa pensare alla possibilità di un meccanismo di autoregolazione di espulsione. Il gas può essere accelerato a qualsiasi velocità necessaria per espellerlo permanentemente; in seguito il meccanismo di accelerazione, ultimato il suo compito, non deve più faticare. Si può anche capovolgere il discorso. Se la stella non è compatta, perché dovrebbe produrre e mantenere l'enorme velocità di espulsione che è stata osservata, se fosse in grado invece di sbarazzarsi per sempre del materiale impartendogli una velocità molto inferiore?

Se SS 433 ospita una stella di neutroni, l'orologio da 164 giorni (cioè il meccanismo che fa ruotare l'asse del getto) probabilmente non è semplicemente la rotazione della stella ogni 164 giorni. Ciò perché le piccolissime dimensioni della stella impongono che a tale bassa velocità di rotazione l'oggetto non sia molto rigido. (Tecnicamente si dice che il suo momento d'inerzia non è molto grande.) Del resto, il materiale espulso trasporta via una tremenda quantità di energia a una velocità fantastica. Il rinculo risultante conferito alla stella cancellerebbe molto rapidamente il comportamento periodico a meno che le due masse di materiale espulso siano esattamente allineate e abbiano la stessa velocità, in modo tale che i loro effetti sulla stella si annullino perfettamente a vicenda. Pare più probabile che il periodo di 164 giorni sia invece un effetto di precessione, un lento rollio dell'asse di rotazione. L'effettivo periodo di rotazione della stella potrebbe in tal caso essere rapido, come nel caso dei radiopulsar, dove i periodi sono dell'ordine di pochi secondi o inferiori. Una valida analogia è un giroscopio per bambini, la cui ruota può ruotare molto rapidamente, mentre tutto l'insieme descrive lentamente un cerchio.

Qual è la sorgente del materiale espulso? Dal momento che le righe di emissione appartengono all'idrogeno e all'elio, elementi presenti normalmente in oggetti molto meno evoluti delle stelle di neutroni, il materiale non viene probabilmente da tale stella. Il mezzo interstellare è composto principalmente da idrogeno ed elio, ma è troppo tenue per fornire gas a sufficienza. Si è pertanto indotti a sospettare la presenza di una seconda stella meno evoluta.

Questa congettura portò David Crampton, Anne P. Cowley e John B. Hutchings del Dominion Astrophysical Observatory in Canada a ricavare una serie di spettrogrammi di SS 433 che potrebbero rivelare tale stella compagna. Essi scoprirono ben presto che anche le righe di emissione «stazionarie», cioè le emissioni dell'idrogeno e dell'elio alle loro lunghezze d'onda di laboratorio subiscono ciclicamente spostamenti delle loro lunghezze d'onda. L'ampiezza dello spostamento è molto piccola e corrisponde a 70 chilometri al secondo (circa lo 0,1 per cento dell'ampiezza dello spostamento delle righe mobili), e pertanto esso non poteva venire rivelato dall'apparecchiatura usata da me e dai miei colleghi per le nostre osservazioni. Il periodo della variazione minore è di 13 giorni. Ancora una volta il periodo e l'ampiezza della variazione di velocità osservati condussero a stime delle masse delle due stelle. Una soluzione coerente, pur se non univoca, si ottiene supponendo che entrambe le stelle abbiano una massa paragonabile a quella del Sole. I parametri della soluzione orbitale fanno pensare che la compagna della stella di neutroni sia probabilmente una stella normale, con caratteristiche non molto diverse da quelle del Sole. Sfortunatamente un tale oggetto risulterebbe troppo debole sul fondo di radiazione dei getti per poter essere osservato direttamente.

Perché non si creda che tutti i misteri di SS 433 siano oggi chiariti, passerò in rassegna la moltitudine di problemi inquietanti che restano da risolvere. La maggior parte di essi riguarda le condizioni fisiche del materiale espulso. È possibile calcolare alcuni dei parametri del gas emittente; tuttavia, per fare ciò si deve avanzare qualche ipotesi sulla distanza di SS 433 in modo da poter utilizzare l'intensità osservata dei getti di radiazione per calcolare la luminosità intrinseca. Come la maggior parte delle stime delle distanze astronomiche, essa è estremamente incerta. Sulla base di argomentazioni quale l'intensità dell'assorbimento spettrale dovuto alla presenza del gas interstellare, ho stimato la distanza in circa 4000 parsec. (Un parsec è pari a 3258 anni-luce.) La nostra galassia ha un'estensione di circa 30 000 parsec, quindi SS 433, anche se non si trova nelle immediate vicinanze del Sole, non è molto distante.

È ora possibile calcolare la luminosità dei getti. Le poche righe di emissione mobili hanno mostrato di possedere più energia, di circa un fattore 10, di quella che il Sole irradia a tutte le lunghezze d'onda. Le dimensioni del materiale che irradia nel visibile sono di circa 10 miliardi di chilometri, cioè circa 100 volte la distanza tra la Terra e il Sole, ovvero il doppio delle dimensioni dell'intero sistema solare. Il parametro più sorprendente è però la quantità di energia cinetica che si ritiene sia necessaria per accelerare le notevole massa di materiale alla straordinaria velocità osservata. Pur dipendendo il valore numerico notevolmente da ipotesi incerte, esso è dell'ordine di 10^{39} erg al secondo, pari a un milione di volte l'energia irraggiata in ogni secondo dal Sole a tutte le lunghezze d'onda. Qual è la sorgente di questa fantastica produzione di energia? Non ne siamo affatto sicuri.

Più sono dettagliate le domande che ci poniamo sui getti, più profonda si rivela la nostra ignoranza. Quale meccanismo fornisce questa energia per accelerare il gas, producendo una sola velocità rigorosamente controllata e immutabile? Perché il gas è così freddo in confronto alle temperature previste per tali velocità? Quale processo collima e indirizza il materiale in un getto? L'ultima domanda è particolarmente tormentosa. Disponiamo oggi di una misura sperimentale della larghezza dei getti basata sulla larghezza delle righe spettrali mobili. Se i getti fossero larghi, in un dato istante una striscia così larga presenterebbe una varietà di angoli così come velocità differenti e spostamenti Doppler differenti. Ci aspetteremmo pertanto che le righe mobili siano larghe, fino a coprire tale intervallo di velocità Doppler. In realtà osserviamo l'opposto: le righe mobili sono molto strette in confronto con le enormi velocità dei getti. La conclusione è che i getti sono larghi meno di qualche grado, quasi come due aghi appuntiti.

Infine si pone questa domanda, forse la più complessa: dove sono gli altri oggetti come SS 433? Perché in una galassia di 100 miliardi di stelle osserviamo solo uno di tali oggetti? Una possibile risposta è che la vita media di questo curioso evento possa essere molto breve in confronto ai consueti valori astronomici, forse solo dell'ordine di 10 000 anni. Molte stelle possono attraversare questa fase, ma solo per un istante astronomico. Può darsi che ci sia soltanto uno di tali oggetti attivo in un istante qualsiasi.

È certamente stupido ipotizzare ciò che si potrà dire sul significato generale per l'astrofisica di SS 433 tra cinque o 10 anni. Ciò nonostante, è interessante pensarci. La possibilità più interessante è che la somiglianza tra la struttura a doppio getto di SS 433 e l'emissione radio a due lobi dalle galassie giganti e dai quasar non sia una coincidenza. Se lo stesso meccanismo fondamentale è alla base di entrambi i fenomeni e se le risposte alle domande di cui sopra relative all'energetica, all'accelerazione e alla collimazione sono correlate nei due casi, sarebbe estremamente affascinante. In tal caso noi saremmo privilegiati poiché potremmo disporre di una visione rivelatrice di un oggetto relativamente vicino, entro la Galassia, che potrebbe essere utilizzato come campione per acquisire conoscenze sui violenti eventi extragalattici che costituiscono i più grandi misteri della natura.

Gli anticorpi monoclonali

Si possono rendere immortali le cellule che secernono gli anticorpi fondendole con cellule tumorali e clonando gli ibridi. Ogni clone può fornire quantità sostanziali di un singolo anticorpo molto specifico

di Cesar Milstein

Quando una sostanza estranea penetra o viene iniettata nel corpo di un vertebrato, un aspetto della risposta immunitaria di questo animale è la secrezione di anticorpi da parte delle cellule plasmatiche, cioè di molecole di immunoglobuline con siti di combinazione che riconoscono, sulla superficie della sostanza estranea, o antigene, la forma di particolari determinanti e si legano a essi. La combinazione dell'anticorpo con l'antigene mette in moto dei processi che possono neutralizzare ed eliminare la sostanza estranea. A parte la loro funzione naturale nella risposta immunitaria, gli anticorpi sono anche stati per molto tempo un importante strumento per i ricercatori, i quali fanno tesoro della loro specificità nell'identificare o nel marcare particolari molecole o cellule e nel separarle da una miscela.

La risposta di un anticorpo a un antigene tipico è estremamente eterogenea. Nella milza di un topo o di un uomo esiste forse un milione di linee di discendenza diverse di linfociti *B*, precursori delle plasmacellule. Derivano tutte da una cellula staminale comune, ma ciascuna sviluppa una capacità indipendente di produrre un anticorpo in grado di riconoscere un differente determinante antigenico. Quando si inietta un agente immunizzante in un animale, questo reagisce producendo diversi anticorpi, diretti contro differenti molecole antigeniche, presenti nella sostanza iniettata, e contro differenti determinanti, presenti su un singolo antigene. Ma produce anche differenti anticorpi che si adattano, più o meno bene, a un unico determinante. È quasi impossibile separare i vari anticorpi e così gli antisieri convenzionali ne contengono miscele, le quali variano poi da animale ad animale.

Ogni anticorpo viene prodotto, tuttavia, da una differente linea di linfociti e dalle cellule plasmatiche che da essi derivano. Che cosa succederebbe se si riuscisse a isolare una di queste cellule e la si coltivasse in vitro? La progenie di questa singola cellula, o clone, sarebbe la fonte di notevoli quantità di uno stesso anticorpo contro un unico determinante antigenico: un anticorpo monoclonale. Sfortunatamente, però, le cellule produttrici di anticorpi non si possono conservare in un mezzo di coltura.

Vi sono, invece, tumori maligni del sistema immunitario, chiamati mielomi, le cui cellule, che proliferano a grande velocità, producono grandi quantità di immunoglobuline anomale, chiamate proteine del mieloma. Un tumore è, in sé, un clone immortale di cellule derivate da un unico progenitore, per cui le cellule del mieloma possono essere coltivate senza limiti e la struttura chimica di tutte le immunoglobuline che secernono risulta identica. In effetti, queste immunoglobuline sono anticorpi monoclonali, ma non c'è modo di sapere contro quale antigene sono dirette, né si possono indurre mielomi che producano un anticorpo contro uno specifico antigene.

Nel 1975, i miei colleghi e io abbiamo imparato il modo di fondere le cellule di mieloma di topo con i linfociti della milza di topo immunizzato con un particolare antigene. Le cellule del risultante mieloma ibrido, o «ibridoma», esprimono sia la proprietà dei linfociti di produrre un anticorpo specifico sia il carattere di immortalità delle cellule del mieloma. Tali cellule ibride possono venir manipolate con tecniche applicabili a cellule animali in coltura permanente. Le cellule ibride singole possono essere clonate e ogni clone produce forti quantitativi di anticorpi identici, diretti contro un unico determinante antigenico. I cloni possono essere mantenuti in vita all'infinito e, in qualsiasi momento, si possono coltivare in vitro campioni prelevati da essi, oppure si possono iniettare in animali per ricavarne anticorpi monoclonali su vasta scala. Anticorpi monoclonali estremamente specifici, prodotti grazie a questo metodo generale, sono risultati uno strumento molto versatile in molte aree della ricerca biologica e della medicina clinica.

I mielomi umani sono noti ai medici da molto tempo, ma la natura esatta delle loro proteine è stata determinata dagli immunologi soltanto agli inizi degli anni sessanta. In seguito, Michael Potter del National Cancer Institute ha indotto dei mielomi in topi e anche questi mielomi hanno prodotto notevoli quantità di immunoglobuline monoclonali. Tuttavia, malgrado i notevoli sforzi compiuti, non è stato possibile indurre tumori che potessero sintetizzare anticorpi contro un antigene iniettato. Presso il Salk Institute for Biological Studies, Leo Sachs, Kenko Horibata, Edwin S. Lennox e Melvin Cohn sono riusciti a ottenere in coltura una linea di cellule di mieloma di topo, ma in seguito essa è andata perduta. Alla fine, Horibata e A. W. Harris sono comunque riusciti a coltivarne una nuova serie, distribuendola ad altri laboratori. Il mio gruppo che lavora presso il Medical Research Council Laboratory of Molecular Biology di Cambridge ha studiato in maniera approfondita una linea ottenuta da uno dei tumori di Potter.

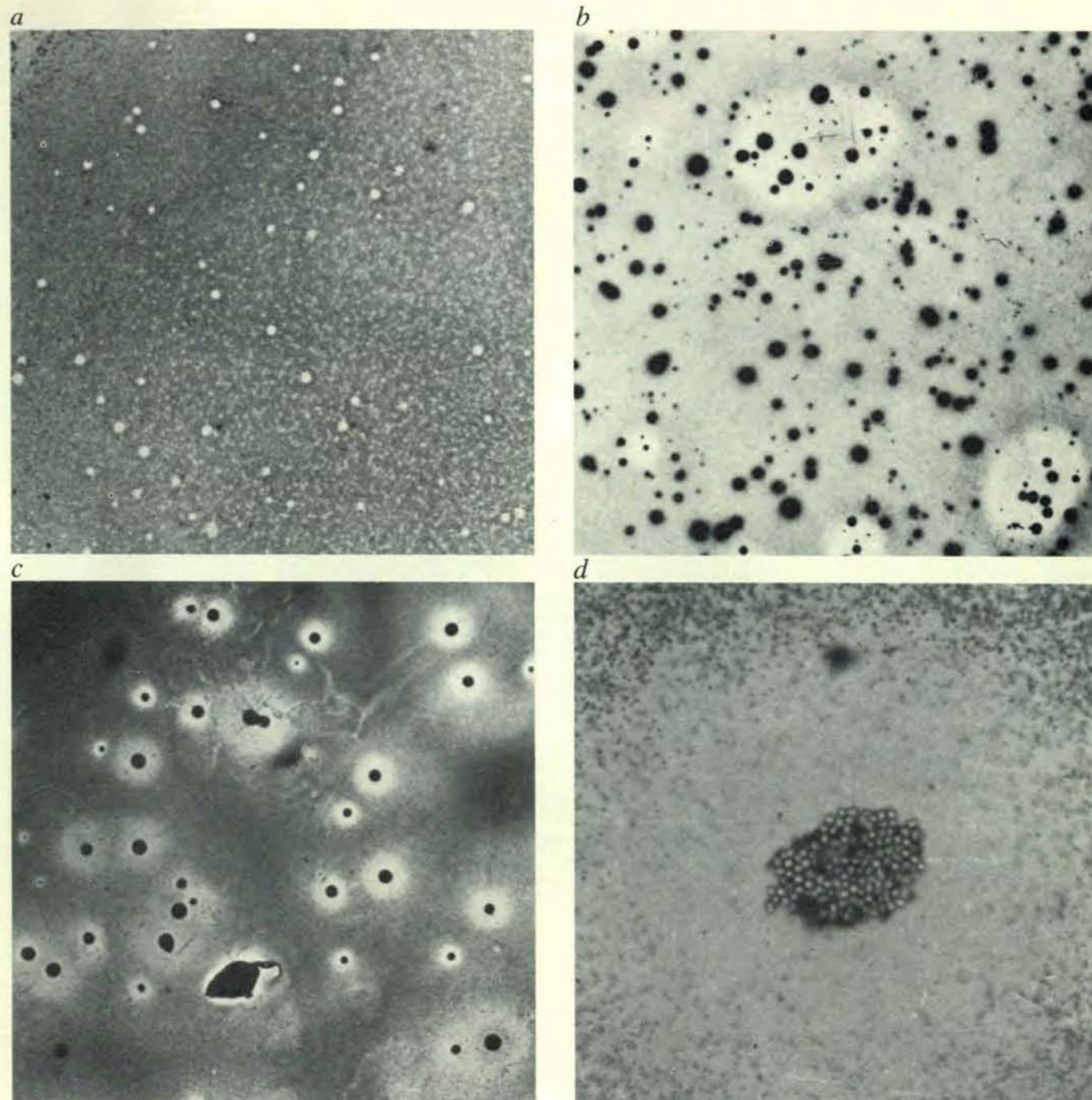
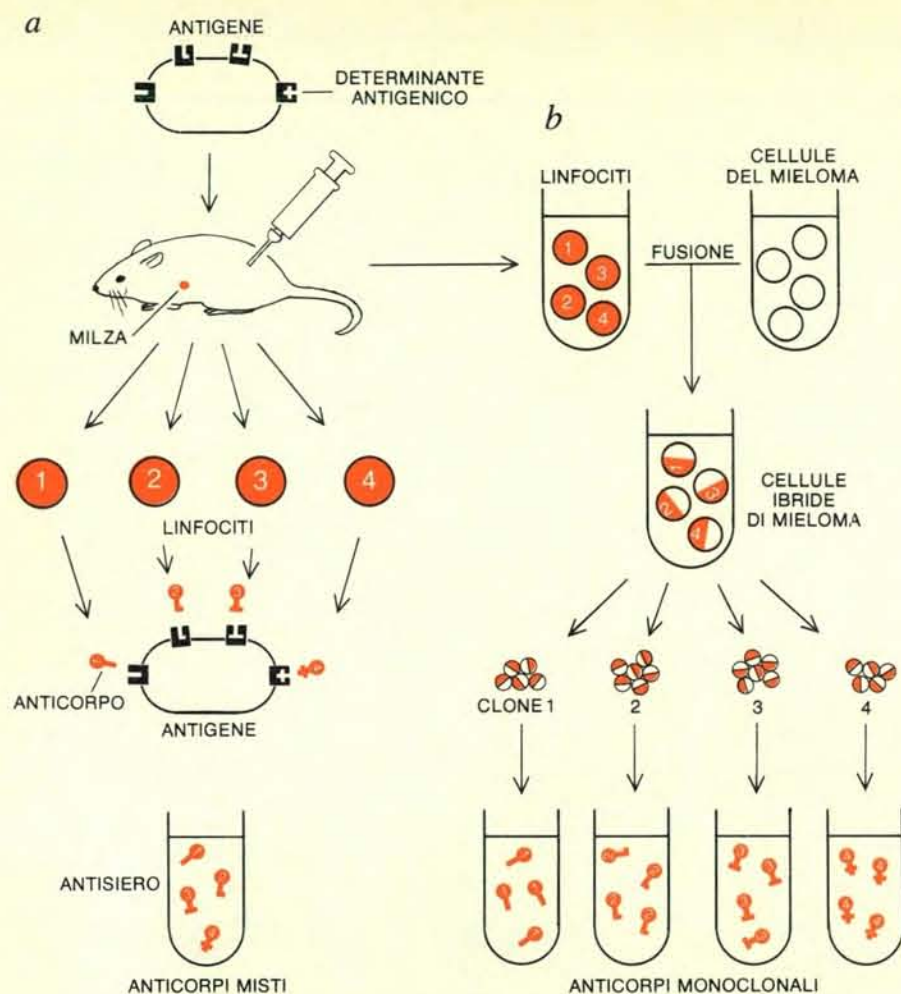
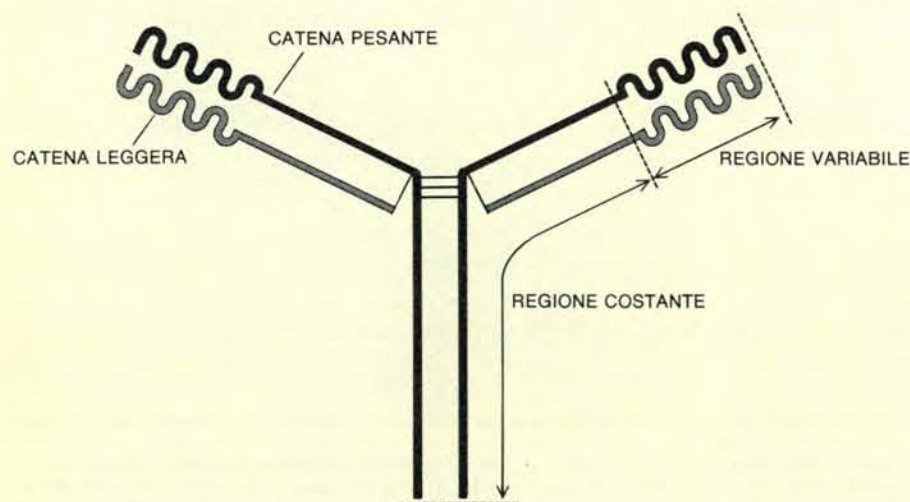
A quell'epoca non pensavamo ancora agli anticorpi monoclonali. Stavamo studiando in che modo le cellule somatiche (corporee) si diversificano in coltura e in che modo le mutazioni modificano la specificità combinatoria degli anticorpi; la linea cellulare del mieloma di topo ci serviva dunque semplicemente come un tipo adatto di cellule da coltivare in vitro. Nel 1973, per la prima volta, Richard G. Cotton, David S. Secher e io siamo riusciti a produrre mutanti strutturali di una proteina di mieloma di topo, secreta da una linea cellulare coltivata. Questo lavoro e indagini parallele condotte da Matthew D. Scharff dell'Albert Einstein College of Medicine di New York hanno dimostrato che mutazioni spontanee nelle cellule coltivate influenzavano la struttura delle proteine che tali cellule sintetizzavano, fornendo inoltre informazioni sulla natura molecolare delle mutazioni e sulla loro frequenza. La ricerca dei mutanti è stata, tuttavia, laboriosa perché le proteine prodotte dalle cellule parentali non avevano un'attività anticorpale manifesta, cioè dei cambiamenti in cui vi fosse l'indicazione più efficace delle lievi differenze provocate dalle mutazioni. È chiaro che ciò di cui si aveva bisogno era una linea cellulare in grado di secernere un'immunoglobulina con una attività anticorpale facilmente analizzabile. Una simile linea non esisteva.

A questo punto una fortunata circostanza ci condusse alla tecnica del mieloma ibrido. Mentre lavoravamo sulle mutazioni somatiche, Georges Köhler e io seguivamo anche un diverso indirizzo di ricerca, nel tentativo di saperne di più sul controllo genetico della sintesi degli anticorpi. Questa viene controllata da due serie di geni. Una codifica la regione «variabile» delle catene leggere e pesanti dell'anticorpo, cioè quella regione che controlla la specificità della molecola dell'anticorpo; l'altra codifica la regione «costante» di tali catene, cioè quella regione che è responsabile di funzioni effettrici come la fissazione del complemento (un complesso di proteine del plasma sanguigno, interessate nella risposta immunitaria), il trasporto della molecola anticorpale attraverso le membrane e la sua fissazione alle membrane stesse. Ogni linfocita sintetizza un anticorpo codificato da un'unica coppia di geni *V* (variabile) e *C* (costante) in mezzo a una vasto reperto-

a

b

c

d

I cloni secernenti anticorpi, costituiti da cellule di mieloma ibrido, sono stati individuati la prima volta grazie a un test per gli anticorpi contro i globuli rossi del sangue di pecora. Nella prova standard *(in alto a sinistra)* i globuli rossi e le cellule che secernono gli anticorpi vengono incubati su agar e viene aggiunto il complemento (un complesso proteico derivante dal plasma sanguigno). Gli anticorpi che diffondono da ciascuna cellula secernente si legano agli antigeni presenti sulle vicine emazie, dando così inizio a una reazione del complemento che uccide tali cellule, formando una placca, cioè una zona chiara *(macchie bianche)* tutt'attorno alle cellule secernenti. L'autore e Georges Köhler hanno fuso cellule di topo che secernono anticorpi contro i globuli rossi con cellule tumorali (di mieloma) e hanno poi seminato su piastra gli ibridi *(in alto a destra)*. Si sono quindi formate colonie di cellule ibride *(macchie nere)* e quando, assieme al complemento, è stato aggiunto anche uno strato di globuli rossi di pecora, solo poche di queste colonie hanno dato origine a placche *(zone bianche attorno alle colonie)*, indicando così che secernevano un anticorpo specifico. Da una colonia di cellule secernenti anticorpi sono state prelevate singole cellule e sono state deposte su piastra in strato sottile *(in basso a sinistra)*; la maggior parte dei cloni derivati sono stati in grado di secernere l'anticorpo contro i globuli rossi. Nella microfotografia in basso a destra sono visibili le singole cellule del clone e una zona di cellule morte attorno a esso.

Una risposta immunitaria ha inizio *(a)* quando una molecola di antigene penetra nel corpo di un animale. Il sistema immunitario reagisce: le linee cellulari dei linfociti *B* proliferano e ciascuna secerne una molecola di immunoglobulina che si adegua a un singolo determinante antigenico. Un antisiero convenzionale contiene una miscela di questi anticorpi. Gli anticorpi monoclonali si ottengono per fusione di linfociti provenienti dalla milza con cellule di mieloma maligno *(b)*. Le singole cellule ibride vengono clonate e ogni clone secerne un anticorpo monoclonale che si adatta a un singolo determinante antigenico sulla molecola dell'anticorpo.

Gli anticorpi appartengono alla famiglia di proteine nota con il nome di immunoglobuline. La forma fondamentale di una molecola di immunoglobulina è quella della classe delle immunoglobuline G (IgG), una popolazione eterogenea di molecole che hanno tutte una struttura a *Y*, composta da due tipi di catene molecolari, legate da ponti disolfuro. Il numero e l'esatta posizione di tali legami sono variabili e caratteristici della sottoclasse delle IgG. Ogni catena presenta poi due regioni: nella regione variabile, le sequenze amminoacidiche offrono siti di combinazione con una diversa conformazione, che si legano in maniera specifica ai differenti antigeni. La regione costante delle catene è responsabile invece di altre funzioni.

rio di geni della cellula. Quando esistono differenti alleli, o forme varianti, di un gene *V* o *C* su ciascuno di due cromosomi della cellula, solo l'allele su uno dei cromosomi è attivo, mentre l'altro no.

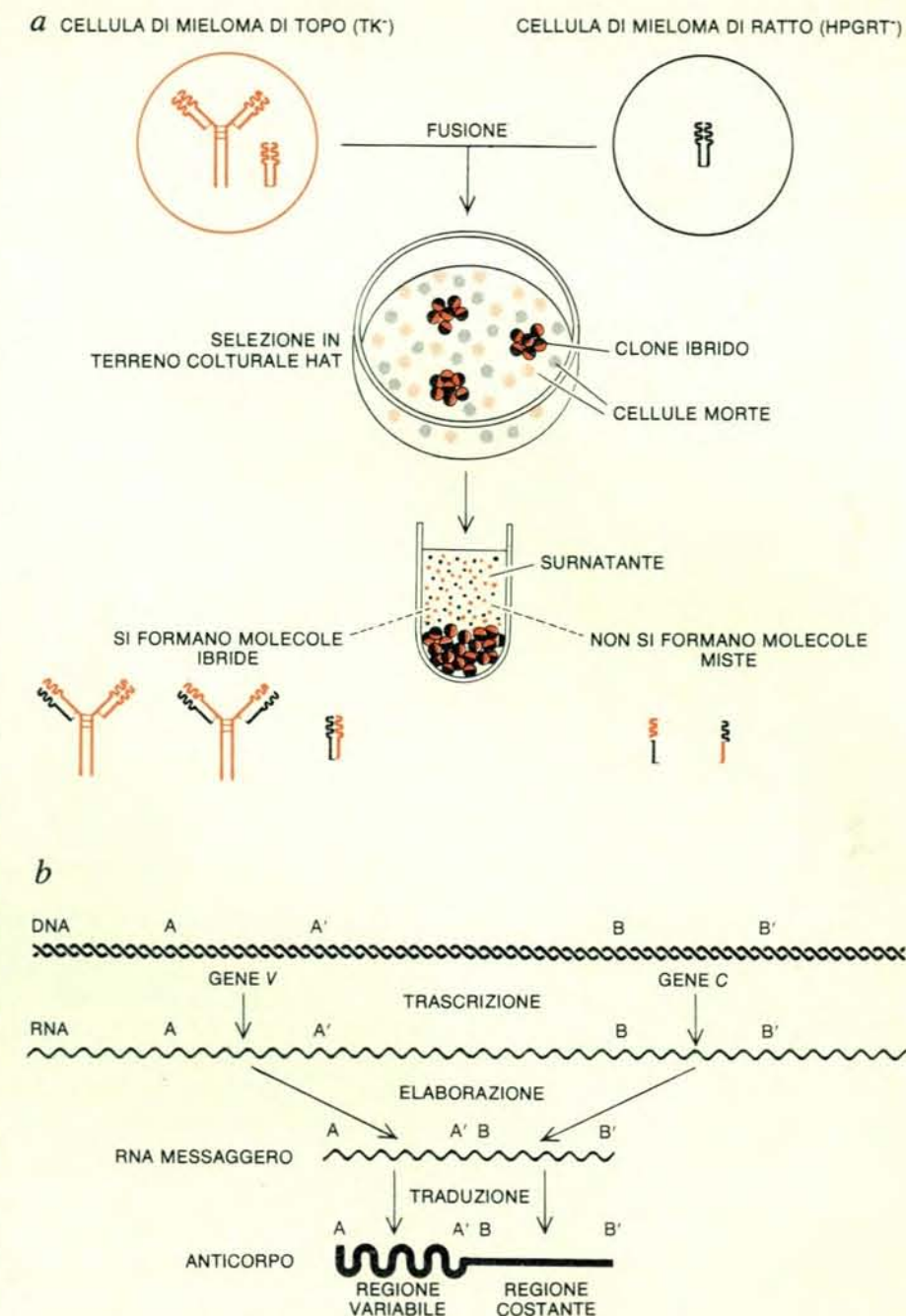
Nel 1973, Cotton e io abbiamo compiuto un esperimento per trovare se si poteva infrangere quest'esclusione allelica e, in caso positivo, quali sarebbero state le conseguenze molecolari. Abbiamo fuso due cellule di mieloma, una proveniente da una linea cellulare di topo e una da una linea cellulare di ratto. L'analisi compiuta sulle cellule ibride ha mostrato che tali cellule secernevano delle molecole pure ibride, consistenti di varie combinazioni di catene sintetizzate dalle cellule parentali, ma mai di una combinazione di una regione *V* di un animale e di una regione *C* dell'altro. Ciò voleva dire che i geni per le regioni *V* e *C* dovevano essere situati sullo stesso cromosoma. Si sa oggi che le sequenze di DNA che codificano per le regioni *V* e *C* sono separate da introni, o sequenze intercalate di DNA. L'intero filamento di DNA viene trascritto in un filamento complementare di RNA nucleare. L'RNA viene poi trasformato da enzimi cellulari (l'introne viene tagliato via e le sequenze che codificano per le regioni *V* e *C* sono saldate assieme) e si forma l'RNA messaggero, che viene quindi tradotto nella sequenza polipeptidica della catena dell'immunoglobulina. In altre parole, l'esperimento con il mieloma ibrido ha mostrato che la saldatura delle sequenze *V* e *C* ha luogo all'interno di un'unica molecola di RNA.

L'esperimento ha anche dimostrato che, negli ibridi, non vi è alcuna esclusione allelica, dato che l'informazione proveniente da ambedue i genitori viene espressa in maniera «codominante» dalle cellule fuse. Questo risultato ha suggerito a Köhler e a me una possibile risposta al nostro bisogno di ottenere, nell'esperimento di mutazione, una cellula produttrice di anticorpi. Ci è capitata l'eventualità di fondere un normale linfocita o una cellula plasmatica con una cellula di mieloma e di immortalare in questo modo l'espressione della secrezione di un anticorpo specifico da parte della cellula plasmatica. Vorremmo ora applicare questa tecnica di fusione cellulare a un nuovo scopo e precisamente usarla per fissare in una linea cellulare permanente una funzione che viene normalmente espressa solo in una cellula «terminale»: la cellula plasmatica derivata da un linfocita *B* stimolato da un antigene.

Per il nostro primo tentativo, abbiamo scelto globuli rossi di pecora come immunogeni, dato che gli anticorpi contro questo tipo di cellule vengono facilmente identificati con una prova messa a punto nel 1963 da Niels Kaj Jerne e da Albert Nordin. In presenza di un agente che promuove la fusione cellulare, abbiamo mescolato cellule di mieloma di topo con cellule della milza di topi immunizzati, abbiamo identificato con successo le cellule fuse in un ambiente selettivo e abbiamo trovato che secernevano immunoglobuline di ambedue i genitori. Alcune secernevano anticorpi contro i globuli rossi *(si veda l'illustrazione a pagina 31)* e siamo riusciti a isolare cloni in grado di produrre singole specie molecolari di quegli anticorpi, cloni che potevano essere conservati in coltura. Per la prima volta avevamo ottenuto colture continue di cellule fuse secernenti un anticorpo monoclonale di specificità predefinita.

Dopo il nostro iniziale successo nel 1975, siamo incorsi in difficoltà e per quasi sei mesi gli esperimenti nel nostro laboratorio o riuscirono male o non riuscirono affatto. Köhler, che si era trasferito all'Istituto di immunologia di Basilea, aveva avuto anch'egli alcune difficoltà. A un certo momento venne nel nostro laboratorio, per lavorarvi, Giovanni Galfré e cercò di apportare al metodo varie modifiche, in particolare l'uso di glicol polietilenico come agente di fusione. Nonostante alcune difficoltà, riuscimmo a ottimizzare le condizioni in ogni fase del processo. Alla fine, fu scoperto il problema di fondo: un lotto tossico di uno dei reagenti. Una volta rimediato il guaio, con Jonathan C. Howard e Geoffrey W. Butcher dell'Agricultural Research Council Institute of Animal Physiology di Cambridge abbiamo realizzato delle spettacolari fusioni, le quali hanno prodotto una serie di anticorpi monoclonali contro gli antigeni di istocompatibilità del ratto: i marcatori della superficie cellulare, che stabiliscono l'identità individuale e sono responsabili del rigetto dei trapianti. Quando siamo riusciti a stabilire un protocollo standard per gli esperimenti e abbiamo elaborato nuovi metodi di saggio della secrezione di anticorpi, altri risultati hanno cominciato a emergere a un ritmo veloce.

Il successo raggiunto con tali esperimenti è stato favorito da un carattere inatteso. Delle cellule di milza di cui ci servivamo per la fusione, forse solo una su cento era una cellula plasmatica che secerneva attivamente anticorpi, mentre d'altra parte circa uno su dieci dei nostri cloni ibridi produceva anticorpi. Cioè si avevano dieci volte più ibridi immortali, positivi, di quanti ci si aspettasse se l'immortalità veniva casualmente trasferita alla popolazione eterogenea di cellule della milza: chiaramente, assieme all'immortalità si era realizzata una selettività. La spiegazione per quest'ultima non è del tutto chiara, ma in base a fatti recenti essa consta probabilmente di due componenti. Da una parte, sembra che la secrezione venga amplificata per cui i linfociti sintetizzano l'anticorpo, ma normalmente non lo secernono, dando luogo a ibridi che, invece, sia lo sintetizzano sia lo secernono. Probabilmente il mieloma, per parte sua, fornisce il meccanismo per la secrezione a qualche produttore di anticorpi che ne è privo. D'altra parte, le condizioni in cui si svolge la fusione rendono chiaramente improbabile che cellule della milza diverse dai linfociti *B* possano dare origine a ibridi dalla vita lunga.

Quando si è ben costituito un clone di cellule fuse, per definizione tutti gli anticorpi che esso secerne derivano geneticamente da un'unica cellula. Tuttavia, non è necessario che si tratti di anticorpi monoclonali nel senso immunologico del termine, dato che ogni cellula del clone ibrido presenta alcuni cromosomi derivanti dalla cellula parentale del mieloma e altri dalla cellula parentale della milza ed esprime ambedue le serie di cromosomi. Potenzialmente una cellula ibrida, invece di produrre soltanto i due componenti di un vero anticorpo monoclonale (un tipo di catena pesante e un tipo di catena leggera), può produrre due catene pesanti e due catene leggere. Chiamiamo una simile cellula HLGK dato che secerne le catene pesanti e leggere della cellula parentale della milza e le catene corri-

La prima fusione di cellule di mieloma, che si è conclusa con successo, è avvenuta tra una linea cellulare di topo che secerneva molecole di IgG complete e un eccesso di catene leggere e una linea cellulare di ratto che secerneva soltanto catene leggere. I mutanti di ciascuna linea, rispettivamente carenti degli enzimi TK e HPGRT, sono stati coltivati insieme su un terreno selettivo, HAT, in cui sono necessari per la sopravvivenza cellulare entrambi i suddetti enzimi; solo le cellule fuse sopravvivevano, formando dei cloni ibridi. Questi cloni hanno secreto nel surnatante varie molecole ibride consistenti di catene pesanti di topo e di catene leggere di ratto *(a sinistra)*, mentre non hanno prodotto molecole miste con regioni variabili di un animale e regioni costanti dell'altro animale *(a destra)*. Questi risultati hanno messo in luce che le regioni variabili e quelle costanti sono trascritte dal DNA nell'RNA e l'RNA viene poi elaborato e tradotto in proteina *(b)*. Hanno dimostrato, inoltre, che l'informazione genetica di ambedue le cellule parentali viene espressa dagli ibridi in maniera «codominante». Studi più recenti sul DNA mostrano che i geni *V* e *C* sono interrotti da sequenze interferenti, che vengono poi eliminate nel corso dell'elaborazione.

spondenti gamma e kappa della cellula parentale derivata dal mieloma. È nella natura stessa delle cellule ibride, particolarmente negli stadi precoci della proliferazione, perdere rapidamente i cromosomi. In questo caso, la perdita ben riconoscibile non è casuale: si perdono generalmente per prime le catene pesanti (H o G) e quindi una o l'altra catena leggera (L o K). Pertanto l'ibrido HLGK dà origine a varianti, che secernono HLK o GLK; queste a loro volta danno origine a varianti che secernono HL, HK, LK, L e K.

Quello che si cerca è il clone di cellule HL, che esprime solo la catena pesante e la catena leggera della cellula splenica dotata di immunità specifica (anche se vi sono valide ragioni per cercare di conservare anche altre varianti, in particolare il clone HK). Clonando, è dunque necessario non solo saggiare l'anticorpo specifico, ma anche analizzare il tipo di catena dell'immunoglobulina e selezionare un clone che secerna decisamente le catene HL (o HK). Il processo selettivo può essere semplificato scegliendo una linea di cellule mutanti del mieloma che esprima soltanto la catena leggera (K) e dia degli ibridi HLK o una linea che non esprima affatto un'immunoglobulina e pertanto dia luogo a un ibrido HL monoclonale dal punto di vista immunologico.

Una volta ottenuto il clone desiderato, si può congelarlo e conservarlo così per lungo tempo. In un qualsiasi momento, se ne può iniettare un campione in animali dello stesso ceppo di quelli che avevano fornito le cellule originali per la fusione. In essi si sviluppano così dei tumori che secernono lo specifico anticorpo monoclonale prodotto dal clone, anticorpo che è presente nel siero in concentrazioni estremamente elevate: generalmente più di 10 milligrammi di anticorpo per millilitro di siero, il che in alcuni casi è equivalente a un titolo di forse un milione. In alternativa, un campione del clone può essere coltivato in massa e l'anticorpo raccolto dal mezzo.

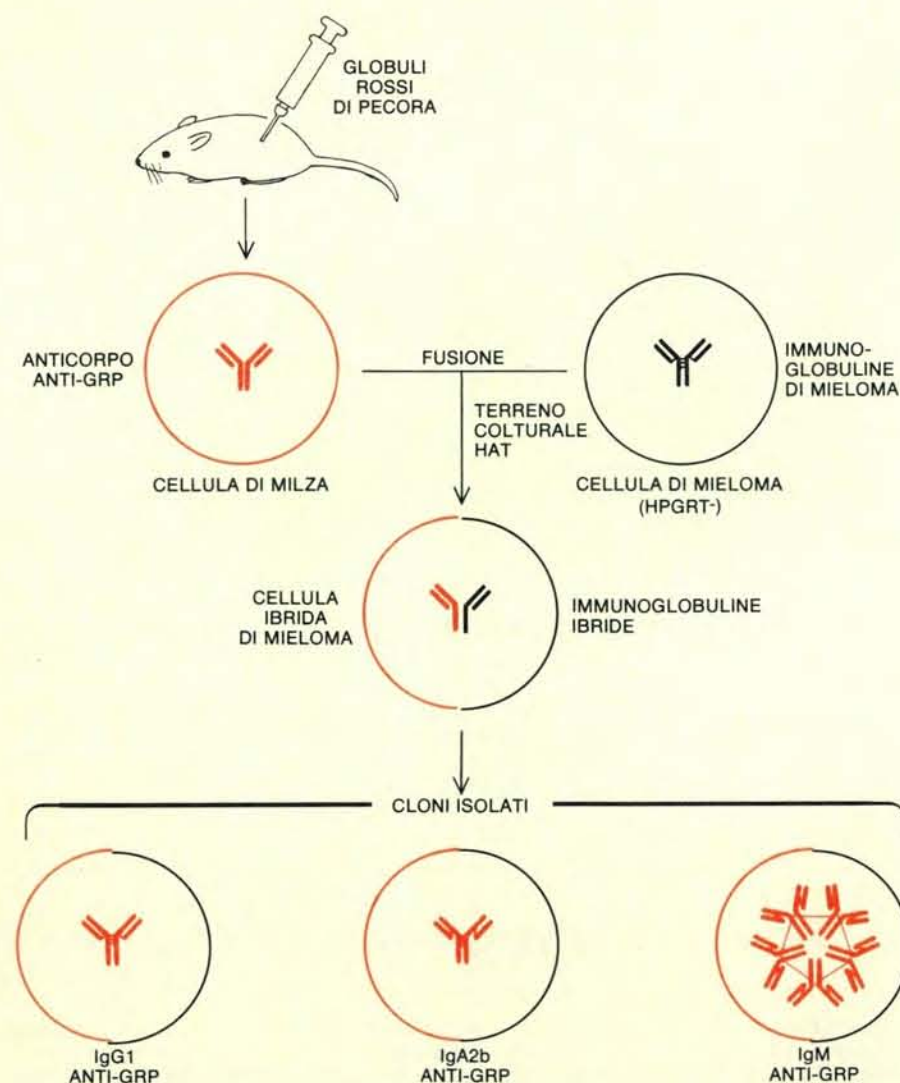
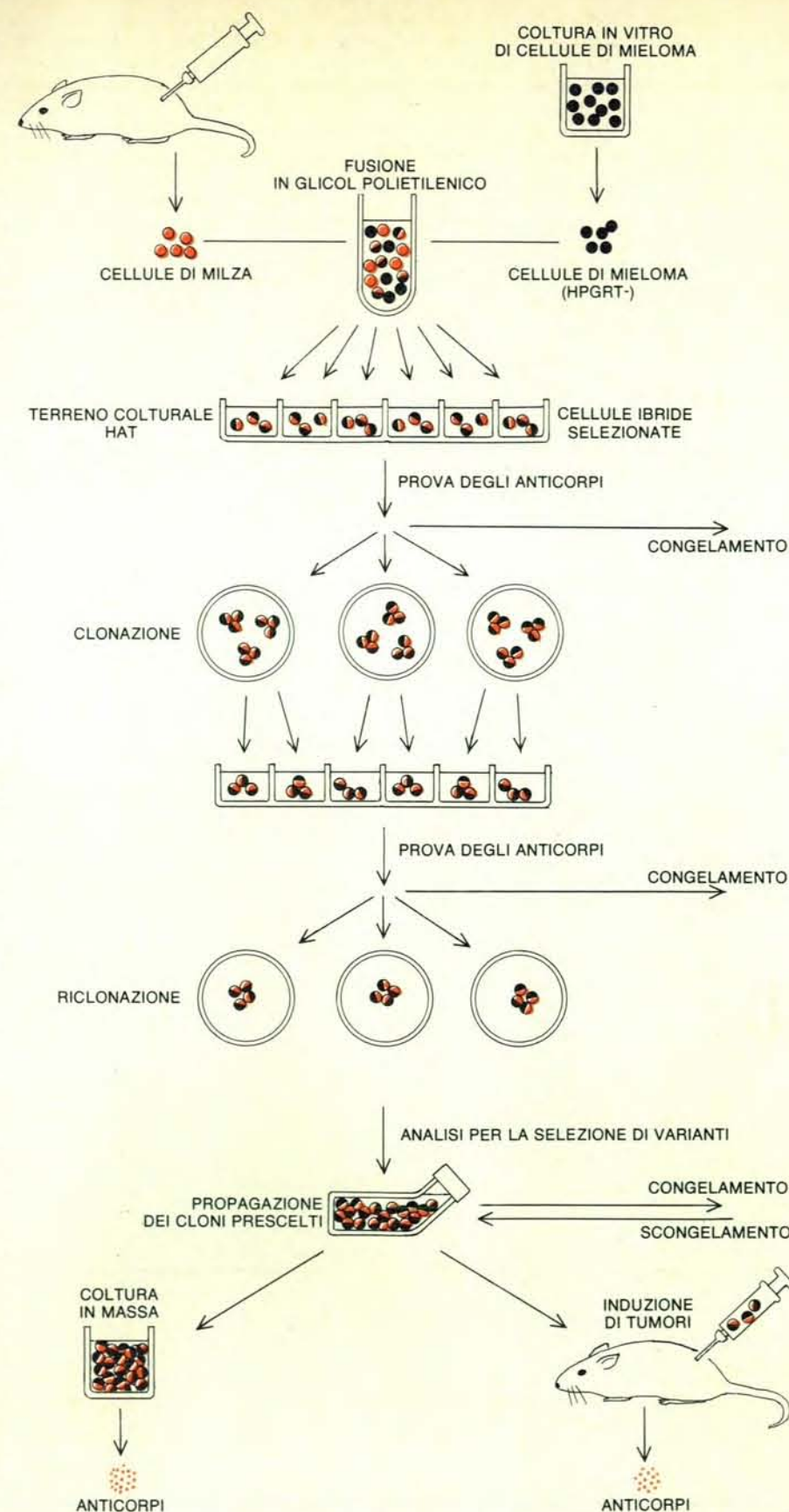
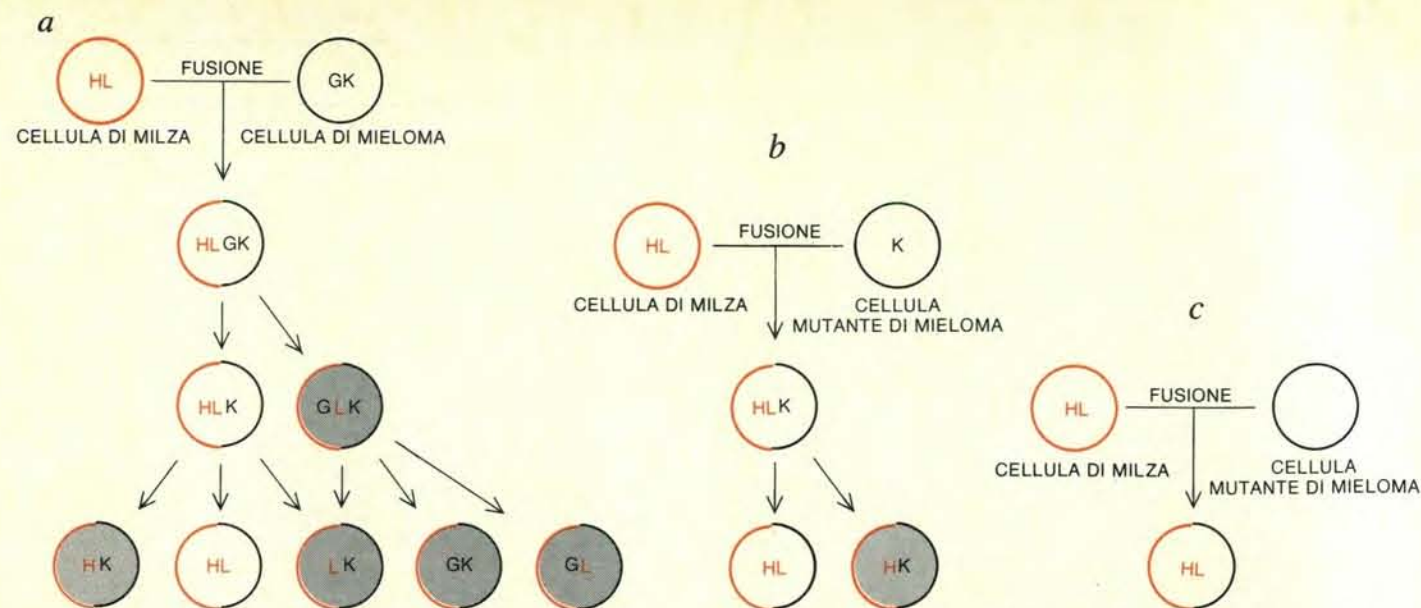
Abbiamo sfruttato metodi simili a quello che ho descritto per produrre anticorpi contro una vasta gamma di sostanze: i piccoli antigeni detti apteni; proteine (tra cui enzimi), carboidrati e glicolipidi; componenti della superficie cellulare e virus. Essi sembrano applicabili in generale. Se l'animale produce un certo anticorpo, si dovrebbe essere in grado di immortalarlo come mieloma ibrido. Il grado di difficoltà nel derivare un mieloma ibrido specifico sembra variare con la risposta immunitaria dell'animale. Quando tale risposta è molto debole, la ricerca di un clone ibrido che secerna l'anticorpo specifico tra i molti cloni che secernono immunoglobuline non specifiche può richiedere speciali procedimenti. Abbiamo messo a punto molti sistemi per selezionare grandi quantità di cellule o di cloni al fine di produrre anticorpi; abbiamo anche imparato il modo di preselezionare milze più ricche di cellule che producono un anticorpo specifico. Tuttavia, non si tratta di miracoli. Se l'animale non produce anticorpi, non c'è modo alcuno per immortalarli.

Un anticorpo monoclonale è un reagente chimico ben definito, che può essere riprodotto a volontà, contrariamente a un antisiero convenzionale, che è un miscuglio variabile di reagenti e non può mai essere riprodotto una volta che sia stata esaurita la fonte originaria. Diventando dunque disponibili gli anticorpi monoclonali, è probabile che essi soppianteranno quelli convenzionali in molti laboratori clinici e di ricerca. Un esempio è dato dal test standard per i gruppi sanguigni *A*, *B*, *AB* e *O*. I reagenti per tale test (gli anticorpi contro gli antigeni *A* e *B* dei globuli rossi) si ottengono da siero umano. I migliori risultati si raggiungono iperimmunizzando il donatore del siero, inoculandogli i globuli rossi del gruppo appropriato (o meglio ancora, un antigene purificato), metodo questo potenzialmente pericoloso; laddove questa iperimmunizzazione non viene effettuata, come avviene nel Regno Unito, i reagenti tendono a essere di qualità inferiore. Inoltre, il siero del donatore deve essere attentamente vagliato per scoprirvi l'eventuale presenza di anticorpi indesiderati, la cui attività potrebbe oscurare la reazione anti-*A* o anti-*B*. (È questo il motivo per cui i reagenti per il test non si possono ottenere immunizzando animali di laboratorio. Gli anticorpi nel siero animale riconoscerebbero il carattere umano dei globuli rossi da saggiare, eliminando completamente la distinzione tra i gruppi *A* e *B*.)

Siamo riusciti a stabilire che un anticorpo monoclonale contro la specificità individuale di un determinato gruppo sanguigno non deve avere origine umana. In collaborazione con Lennox, Steven Sacks e altri viene oggi prodotto un reagente presso il Medical Research Council Laboratory of Molecular Biology, che si ottiene da colture in massa di cellule di mieloma ibrido che riconosce in maniera specifica l'antigene di gruppo *A*. Esso è stato confrontato con i migliori reagenti disponibili in commercio da Douglas Voak e da Jack Darnborough del Cambridge Regional Blood Transfusion Center e si è trovato che ha un ugual grado di efficacia.

La possibilità di ottenere da una miscela «impura» degli anticorpi contro un singolo componente schiude una nuova via di avvicinamento alla purificazione dei prodotti naturali. Convinti di questa premessa, abbiamo sentito la necessità di metterla alla prova nel più rigoroso test al quale potevamo pensare. La sostanza prescelta è stata l'interferon, notoriamente difficile da purificare e da ottenere in quantità significative. Quando Secher e Derek C. Burke della Università di Warwick si sono messi all'opera per purificarlo ricorrendo all'immunoadsorbimento, i migliori preparati che avevano a disposizione erano puri per circa l'1 per cento. Essi hanno immunizzato dei topi con un preparato di questo interferon non purificato, hanno fuso cellule di milza di questi topi immunizzati con cellule di mieloma di topo e quindi hanno saggiato i cloni ibridi per vedere la produzione di anticorpi anti-interferon, arricchendo così con test per la secrezione delle immunoglobuline il saggio biologico laborioso e non attendibile per l'attività anti-interferon. Con notevole difficoltà essi sono stati in grado di selezionare un clone positivo e, iniettando cellule di questo clone in un topo, hanno indotto la formazione di tumori che secernevano l'anticorpo in grandi quantità. Per preparare una colonna immunoadsorbente, hanno immobilizzato questo anticorpo su sferette costituite da carboidrati. Facendo passare poi attraverso la colonna una soluzione assolutamente non purificata di interferon, hanno ottenuto, in una volta sola, una purificazione di 5000 volte. Si sta ora esaminando la possibilità di una purificazione su scala industriale.

Si possono preparare degli anticorpi monoclonali che siano specifici per singoli componenti di una qualsiasi miscela complessa, mentre si possono produrre per le colonne immunoadsorbenti, quantità illimitate di ogni anticorpo. Ciò permette di separare una miscela di sostanze del tutto ignote nei suoi componenti. Si immunizzano animali con la miscela da analizzare; si ricavano cloni di mieloma ibrido e gli anticorpi di ogni clone servono a rimuovere dalla miscela, uno alla volta, con un metodo «a cascata», i singoli componenti.

Una delle più emozionanti applicazioni degli anticorpi monoclonali nella ricerca rientra nel campo della biologia della membrana. Le proteine della membrana sono difficili da purificare. Esse sono presenti nelle cellule in piccole quantità e spesso hanno una attività biologica non facilmente misurabile o,

Le prime cellule di mieloma ibrido sono state preparate immunizzando topi con globuli rossi di pecora (GRP). Vengono qui illustrate in forma schematica le cellule interessate nell'esperimento e le loro immunoglobuline. Cellule di milza di topi immunizzati (che muoiono in qualsiasi coltura in vitro) sono state fuse con cellule di mieloma di topo, carenti dell'enzima HPGRT (che muoiono nel terreno selettivo HAT). Le cellule ibride sono sopravvissute nel terreno HAT e hanno secreto immunoglobuline derivate da ambedue i genitori; alcune poi hanno secreto l'anticorpo attivo contro i GRP (*si veda la figura a pagina 31*). Sono stati isolati cloni di cellule ibride che secernevano solo una singola specie di anticorpi contro i GRP: l'IgG1, l'IgA2b e l'IgM (un pentamero). In altre parole, ogni clone secerneva un anticorpo contro i globuli rossi di pecora.

Il procedimento standard per ottenere anticorpi monoclonali ha inizio con la fusione, mediata da glicol polietilenico, di cellule di milza di topo (o ratto) immunizzato con cellule di mieloma di topo (o di ratto). Gli ibridi vengono poi selezionati su terreno HAT, che viene saggiato per mettere in evidenza la secrezione di anticorpi. Una porzione di ogni coltura positiva viene congelata per precauzione. Le colture positive sono poi clonate e i singoli cloni saggiati. Quelli positivi vengono congelati, riclonati e saggiati per riscontrarvi la presenza di varianti di immunoglobulina (*si veda la figura nella pagina successiva*). I cloni che vengono infine selezionati si possono conservare congelati. Quando poi vengono scongelati, possono o essere coltivati per produrre anticorpi o inoculati in animali per indurvi mielomi che secernono l'anticorpo.

Una cellula di mieloma ibrido, prodotta con il normale processo di fusione *(a)*, sintetizza non solo le catene pesante *(H)* e leggera *(L)* della cellula parentale splenica, ma anche le catene pesante gamma *(G)* e leggera kappa *(K)* della cellula parentale derivata dal mieloma: viene così designata come HLGK. Gli ibridi tendono a perdere i cromosomi, per cui a mano a mano che il clone cresce alcune catene - come indicato qui - vanno perdute. Per selezionare le varianti desiderate *(cellule senza fondino)* vengono saggiate delle colture fino a quando si riesce a isolare una variante stabile HL, che secerne solo un anticorpo specifico (o, per alcune finalità, una variante HK). La fusione con cellule di mieloma mutanti, che sintetizzano solo una catena K *(b)* o nessuna immunoglobulina *(c)* rende più facile derivare il clone HL voluto.

addirittura, la loro attività viene distrutta quando le membrane sono rese solubili per essere analizzate. Un modo per superare questi problemi consiste nel caratterizzare le molecole della superficie cellulare mediante metodi immunologici, un metodo che è stato fruttuoso nel riconoscimento degli antigeni di superficie, che caratterizzano particolari tipi cellulari in differenti stadi del differenziamento dei tessuti. Gli anticorpi convenzionali contro gli antigeni di superficie sono, però, in genere complessi e non riconoscono molecole singole; sono stati necessari procedimenti elaborati per aggirare questa difficoltà.

Nel 1977, Galfré e io, con Alan F. Williams dell'Università di Oxford, abbiamo dimostrato in che modo la tecnica del mieloma ibrido poteva servire a identificare singoli antigeni di differenziamento. Abbiamo immunizzato un topo con membrane cellulari di timo di ratto. Dopo aver immortalato le cellule di topo che secernevano anticorpi contro i linfociti di timo, siamo riusciti a isolare dei cloni che producevano differenti anticorpi specifici. In quel tentativo, abbiamo definito tre nuove specificità antigeniche, un compito che avrebbe potuto richiedere anni di sofisticata immunologia con i metodi convenzionali. Da allora, è stato preparato un certo numero di altri anticorpi monoclonali contro gli antigeni di cellule di topo, di ratto e umane.

La struttura antigenica della superficie cellulare definisce la linea di discendenza di una cellula e le sottoserie di cellule. Per esempio, i linfociti *B* si possono distinguere dai linfociti *T* (che prendono parte alle reazioni immunitarie della cellula piuttosto che alla secrezione degli anticorpi) in buona parte perché hanno molecole di immunoglobulina sulla loro superficie, mentre i secondi hanno marcatori caratteristici come l'antigene Thy-1. La maggior parte dei marcatori di superficie, tuttavia, non mostra specificità per un unica sottoserie di cellule. Perfino l'immunoglobulina caratteristica della cellula *B* può essere presente in parecchi membri funzionalmente diversi della linea di discendenza delle cellule *B*, dalle cosiddette cellule della memoria (che rispondono alla reimmunizzazione con un antigene al quale l'organismo era stato esposto in precedenza) alle plasmacellule (che secernono gli anticorpi). Ciò che caratterizza un dato stato differenziato è la presenza di un certo insieme di antigeni di superficie e la loro espressione quantitativa.

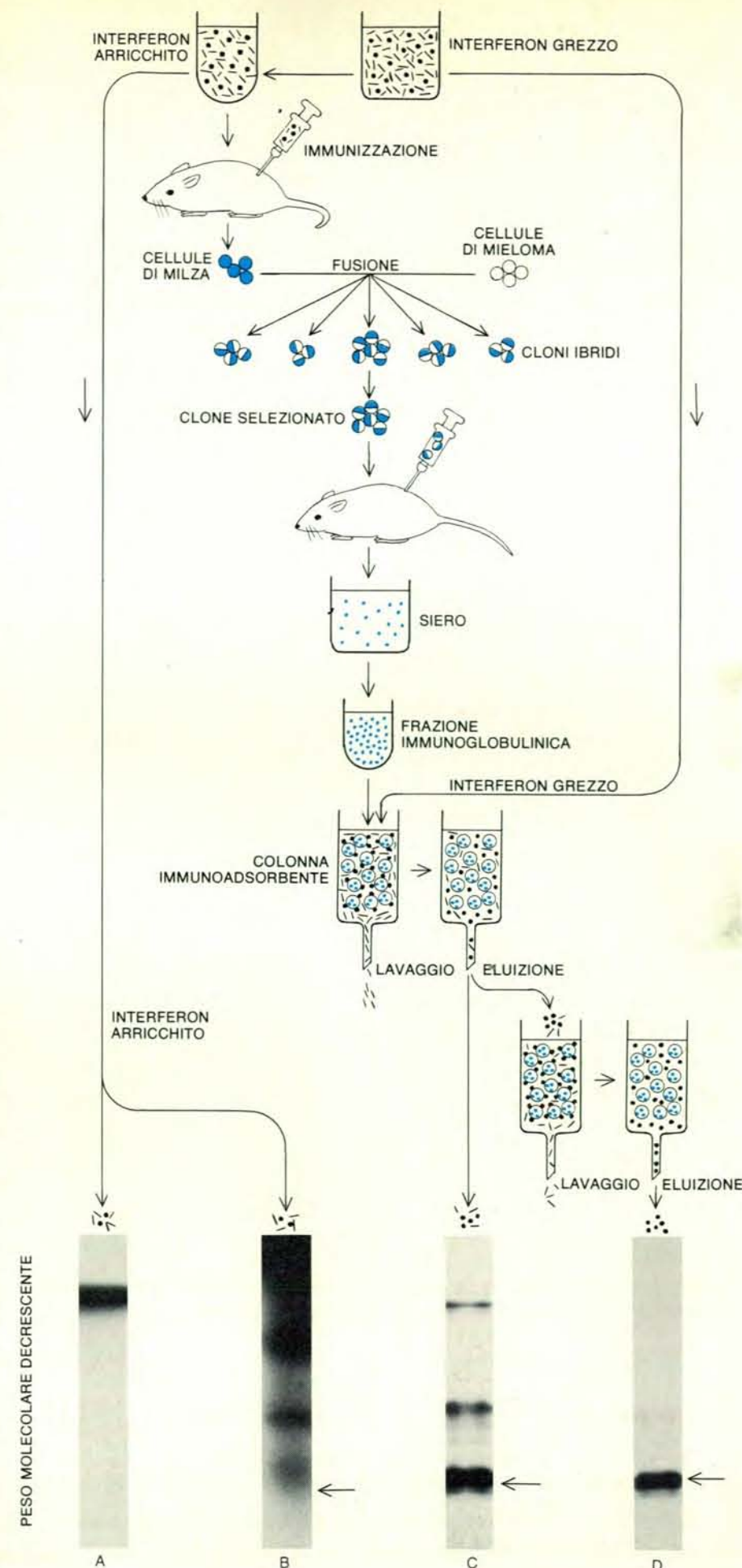
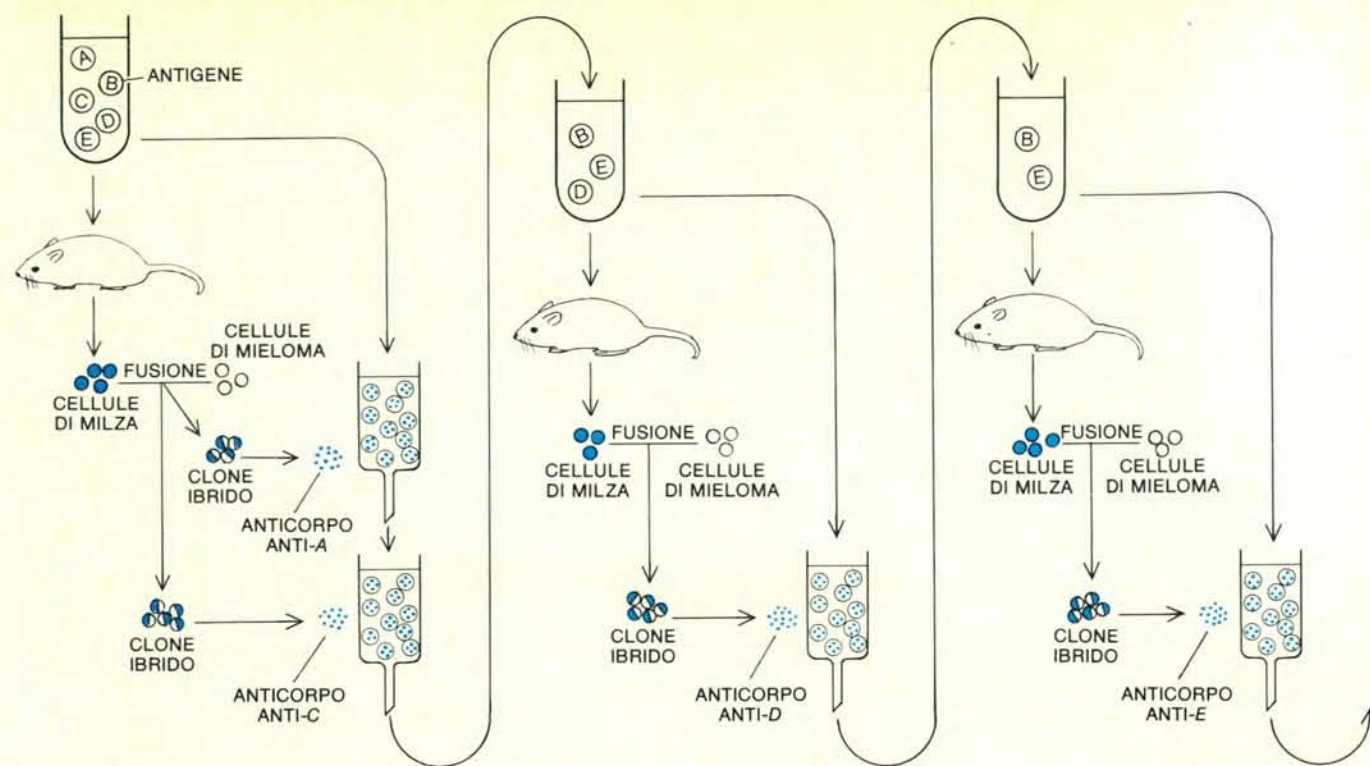
Per stabilire un unico profilo antigenico per ciascuno dei numerosi tipi cellulari sarà indispensabile un'ampia collezione di anticorpi monoclonali e ciò richiederà molto tempo. Un buon inizio si è avuto grazie ai pochi reagenti già disponibili e con l'aiuto di citofluorometri e di selezionatori di cellule, attivati dalla fluorescenza: strumenti che possono rapidamente misurare sia la dimensione sia l'intensità della fluorescenza in grandi quantitativi di cellule a cui siano stati attaccati anticorpi monoclonali, marcati con un colorante fluorescente. Una popolazione cellulare numerosa può così essere frazionata in sottopopolazioni in base alle dimensioni e al quadro dei loro antigeni di superficie; in seguito può essere studiata la funzione di ogni sottopopolazione. In altre parole, gli anticorpi monoclonali sono reagenti standard che possono identificare nuove molecole di superficie e, nel contempo, fare una distinzione tra varie popolazioni cellulari. Finora i migliori risultati sono stati ottenuti con varie cellule ematopoietiche e linfoidi; una applicazione pratica diretta è stata la diagnosi differenziale di varie leucemie e disturbi correlati.

Il quadro della reattività degli anticorpi monoclonali contro le sottopopolazioni di cellule è talvolta conforme con il quadro di maturazione di una data linea cellulare, ma non sempre. Un anticorpo monoclonale sembra riconoscere un antigene caratteristico di certe cellule di midollo osseo nel ratto, mentre negli organi linfoidi periferici riconosce i linfociti e, nel tessuto nervoso, un certo componente che finora non è stato identificato. Nei linfociti periferici, l'antigene è presente sulle cellule *T*, ma non sulle *B*; malgrado ciò ricompare sulle plasmacellule, che pure sono derivate dalle cellule *B*. Noi diciamo che questo anticorpo monoclonale riconosce un tipo di determinante antigenico «saltuario».

Il metodo monoclonale per caratterizzare gli antigeni di differenziamento permette così di sondare il particolare stadio in cui un antigene si esprime, come pure la linea cellulare che lo esprime. Il metodo di purificazione a cascata, che ho descritto sopra, può essere applicato non solo alla caratterizzazione della complessità antigenica della superficie cellulare, ma anche all'analisi dei componenti funzionali e strutturali di altri materiali biologici, come gli organelli cellulari e gli estratti cellulari farmacologicamente attivi.

La «monospecificità» degli anticorpi che si ottengono da cloni di mielomi ibridi ha fatto nuova luce su alcuni fenomeni ben noti di reazioni antigene-anticorpo. Un'indicazione del legame, in provetta, di un antigene con anticorpi convenzionali è, per esempio, la formazione di un precipitato. Questo effetto, in generale, non si osserva quando l'anticorpo è monoclonale. È questa, forse, la prima prova formale della teoria, avanzata più di 40 anni fa, che il precipitato è un reticolo tridimensionale di antigeni e di anticorpi. Un anticorpo monoclonale si lega solo a un unico determinante antigenico, per cui un simile reticolo non può essere formato da un anticorpo monoclonale e dalla maggior parte degli antigeni. Si può formare soltanto se l'antigene è un polimero composto da elementi strutturali identici che si ripetono.

La monospecificità ha anche rivelato alcuni fenomeni fino a questo momento non sospettati, che richiedono nuove interpretazioni delle reazioni antigene-anticorpo. Per fare anche un solo esempio, sembra che il legame di differenti anticorpi con siti vicini dello stesso antigene sia un fattore importante nella rottura della membrana di una cellula da parte del complemento. Abbiamo scoperto questo effetto sinergico quando stavamo isolando gli anticorpi di ratto contro gli antigeni di istocompatibilità. Abbiamo saggiato la presenza di mielomi ibridi secernenti anticorpi, misurando la attività citotossica dei loro terreni colturali. I surnatanti delle colture non clonate erano coerentemente citotossici, ma, clonate le singole cellule, i surnatanti di queste ultime non mostravano affatto una simile attività. È capitato a Howard di misurare l'attività di una miscela di surnatanti di questi cloni chiaramente negativi: con

Un preparato di interferon grezzo è stato purificato mediante immunoadsorbimento. Cellule di milza di topi immunizzati con un preparato in una certa misura arricchito di interferon sono state fuse con cellule di mieloma. Un clone di mieloma ibrido selezionato per l'attività anti-interferon e la secrezione di immunoglobulina è stato inoculato per indurre la formazione di mielomi: l'anticorpo purificato di siero di topi affetti da tumore è stato fissato su sferette di carboidrati per allestire delle colonne immunoadsorbenti. Quando un preparato grezzo di interferon è stato fatto passare attraverso una di queste colonne, si è legato all'anticorpo ed è stato trattenuto mentre altri componenti del miscuglio grezzo sono passati con l'eluente. L'interferon è stato eluito in un secondo tempo. Un solo passaggio attraverso la colonna è sufficiente a fare aumentare l'attività dell'interferon di circa 5000 volte. L'elettroforesi delle proteine (marcate con materiale radioattivo per aumentarne la visibilità) ha reso i risultati visibili *(in basso)*. Il preparato parzialmente arricchito ha mostrato una sola banda *(A)*, chiaramente di albumina; una sovraesposizione dello stesso gel *(B)* ha messo in luce bande minori, una delle quali *(freccia)* in una posizione corrispondente al peso molecolare dell'interferon. Dopo il passaggio di un preparato grezzo di quest'ultimo attraverso una colonna *(C)*, era presente meno contaminante e si è notata anche una banda marcata in corrispondenza dell'interferon *(freccia)*. Un secondo passaggio *(D)* ha dato origine a materiale con una sola banda intensa dell'interferon.

Una miscela ignota può in effetti essere scissa a opera degli anticorpi monoclonali che genera. Una serie casuale di anticorpi, derivata immunizzando animali con una miscela ignota, viene fatta passare su colonne immunoadsorbenti, che rimuovono gli antigeni corrispondenti. La miscela impoverita, ma arricchita di antigeni rimanenti, viene inoculata per produrre più anticorpi con un sistema «a cascata». Le cellule di mieloma ibrido forniscono così uno strumento utile per caratterizzare i componenti e, al tempo stesso, per separarli e purificarli.

nostro compiacimento la miscela è risultata attiva e quindi è stato facile purificare da essa due componenti complementari.

Una volta compreso l'effetto sinergico, l'attività «silente» dei componenti isolati ha potuto essere sfruttata in modo speciale. Le cellule usate come test hanno potuto essere «sensibilizzate», esponendole a un anticorpo monoclonale, e quindi sono state esposte ad anticorpi di altri cloni, rivelando in tal modo interi repertori di anticorpi che agiscono in maniera sinergica. Chiaramente esistono casi in cui, per produrre un effetto desiderato, saranno essenziali miscele di anticorpi monoclonali. In ogni caso si dovrà prendere una decisione e dire se i vantaggi ottenuti dal mescolamento di tali anticorpi in specifiche proporzioni (invece di basarsi su miscele non controllabili quali sono quelle presenti nei comuni antisieri) giustificano lo sforzo impiegato nel ricavarli.

Gli anticorpi monoclonali cominciano lentamente a sostituire gli antisieri convenzionali nell'attrezzatura standard, che viene fornita per procedimenti come il saggio radioimmunologico. Molte ditte li stanno immettendo sul mercato. Dato che possono essere prodotti in grandi quantità permetteranno un vasto impiego di reagenti diagnostici che, fino a questo momento, o non erano disponibili o erano considerati troppo specializzati per un uso generale. Ne è un esempio l'anticorpo contro un neurotrasmettitore chiamato sostanza *P*, ottenuto di recente da A. Claudio Cuello e da me. L'impatto degli anticorpi monoclonali in virologia, parassitologia e batteriologia comincia solo ora a farsi sentire. Grandi speranze sono riposte sulla loro applicazione ai trapianti d'organi, un aspetto dei quali dovrebbe essere la standardizzazione a livello mondiale della tipizzazione dei tessuti. Nella ricerca fondamentale le possibilità sono ancora più ampie e sono già state riferite alcune applicazioni nel campo della embriologia e della farmacologia, come pure nello studio dei recettori per gli ormoni e dei neurotrasmettitori.

Possibili ruoli degli anticorpi monoclonali nella terapia diretta sono in fase di attento esame. La funzione più ovvia sembra essere quella nell'immunizzazione passiva (iniezione di un anticorpo in un paziente, invece dell'immunizzazione attiva in cui un antigene stimola nel paziente una propria risposta anticorpale). Data l'impurità degli anticorpi convenzionali, l'immunizzazione passiva non rappresenta un tipo di trattamento comune, ma può dimostrarsi efficace quando si può somministrare un anticorpo purificato. Nella terapia tumorale si prevedono due ruoli per gli anticorpi monoclonali. Uno di questi consiste nel marcare i farmaci tossici: gli anticorpi contro i tessuti di un particolare organo o contro specifici antigeni tumorali potrebbero essere attaccati alle molecole del farmaco tossico per concentrarne l'effetto. In alternativa, si possono produrre anticorpi antitumorali che troveranno le cellule tumorali, attaccandosi a esse.

Per le applicazioni terapeutiche sarebbero auspicabili anticorpi derivati da linfociti umani invece che da linfociti di topo o di ratto. Ma, contrariamente alle prime speranze, questo fatto si è dimostrato di difficile applicazione; tentativi per immortalare cellule umane produttrici di anticorpi fondendole con cellule di mielomi di topo o di ratto hanno dato finora risultati deludenti. Il problema sta nel fatto che, quando le cellule umane vengono fuse con cellule animali, vi è una rapida perdita preferenziale di cromosomi umani dalle risultanti cellule ibride interspecifiche. E, finora, la ricerca di una linea di cellule di mieloma umano adatta, in grado di essere coltivata in vitro e fusa a dare un ibrido intraspecifico, non ha portato alcun frutto.

In questa panoramica degli impieghi degli anticorpi di mielomi ibridi, ho accennato solo superficialmente alle loro applicazioni più ovvie nella ricerca immunologica fondamentale. Ho preferito sottolineare il fatto che, pur essendo la tecnica nata nel corso del nostro sforzo di capire l'organizzazione e l'espressione genetica delle immunoglobuline, c'è già stata una impressionante «spinta» in molti altri campi. È sempre difficile definire il limite tra ricerca di base e ricerca applicata, ma il fatto di sperimentare personalmente la transizione dall'una all'altra ha fatto su di me una notevole impressione. Non penso che, se lo scopo della mia ricerca cinque o sei anni fa fosse stato quello di produrre anticorpi monoclonali, avrei mai avuto l'idea di tentare simultaneamente di derivare cellule mutanti secernenti anticorpi in un angolo del laboratorio e di fondere due cellule di mieloma in un altro angolo. Eppure questa è stata la combinazione che ha condotto alla produzione iniziale di anticorpi monoclonali contro i globuli rossi di pecora.

Gli antigeni di differenziamento sono determinanti antigenici della superficie cellulare o specifici per i singoli tipi cellulari o comuni a loro serie o sottoserie. Ipotetici antigeni vengono qui illustrati per quattro globuli bianchi. La miglior definizione della genealogia cellulare e delle sottoserie di cellule è il tipo di espressione di questi marcatori, il quale può essere stabilito dagli anticorpi monoclonali.

Questa pietra-camaleonte sembra adattare il suo colore alla luce che la illumina. Si tratta del minerale alessandrite, che appare rosso alla luce di una candela o di una lampada a incandescenza, ricche di radiazioni di grande lunghezza d'onda, mentre diventa verde-azzurro alla luce del giorno o di una lampada fluorescente. L'alessandrite è un alluminato di berillio, $BeAl_2O_4$, ma il suo colore è dovuto alla presenza di impurezze di cromo. Le transizioni fra i diversi livelli energetici degli elettroni spaiati presenti negli ioni cromo sono la causa del colore di molte gemme. Il colore dipende dall'ambiente chimico in cui vengono a trovarsi tali ioni. Piccole variazioni di questo fanno sì che gli ioni cromo colorino in rosso il rubino e in verde lo smeraldo. Nell'alessandrite gli ioni cromo si trovano in condizioni intermedie fra quelle presenti nel rubino e nello smeraldo, con la conseguenza che il colore dei cristalli è molto variabile. Le alessandriti naturali sono molto rare. Il cristallo in figura è stato prodotto artificialmente dalla Creative Crystals Inc. di Concord, California. Le fotografie, che sono state riprese da Fritz Goro, non mostrano gli esatti colori del cristallo perché la sensibilità spettrale delle pellicole a colori è diversa da quella dell'occhio umano.

L'origine dei colori

Pur diversi presentano tutti la medesima origine: sono gli elettroni che rendono il mondo colorato interagendo con le radiazioni visibili di diversa lunghezza d'onda

di Kurt Nassau

Che cosa rende verde un rubino e rosso uno smeraldo? La prima risposta che si può dare alla domanda è molto semplice. Quando la luce bianca attraversa un cristallo di rubino ne emerge modificata, con una percentuale molto alta di radiazioni di grande lunghezza d'onda, che l'occhio vede come luce rossa. La luce che attraversa uno smeraldo ne emerge con una composizione spettrale diversa, che produce la sensazione del colore verde. Si tratta di una spiegazione corretta ma poco soddisfacente. Manca infatti qualsiasi accenno ai meccanismi con i quali la materia è in grado di modificare la composizione spettrale della luce incidente. Il colore del rubino e dello smeraldo sono dovuti alla presenza di una medesima impurezza. Perché possiedono un colore tanto diverso? Qual è l'origine delle sottili differenze che si osservano fra i colori delle diverse sostanze?

Le cause ultime della sensazione del colore sono parecchie e alquanto diverse. Nella classificazione informale, adottata in questo articolo, compaiono 14 categorie di cause, alcune delle quali comprendono più fenomeni fra loro correlati. Ma, con una sola eccezione, i vari meccanismi possiedono un elemento in comune: la causa fisica che produce alla fine la sensazione del colore è l'interazione della radiazione con gli elettroni. Tali interazioni sono state uno dei principali argomenti di ricerca dei fisici del nostro secolo e quindi non può sorprendere che la spiegazione dell'origine dei colori si basi su parecchie teorie fondamentali. In effetti il colore è la manifestazione visibile delle leggi profonde che determinano la struttura della materia.

I gradini dell'energia

La percezione dei colori è un'esperienza soggettiva nella quale hanno un ruolo importante fattori fisiologici e psicologici. Non ci occuperemo in dettaglio di tali problemi. Sembra comunque ragionevole, ai nostri fini, assumere che il colore percepito è il risultato di una misura eseguita dall'occhio e di un'interpretazione data dal cervello della composizione spettrale della radiazione osservata. Il significato di questa assunzione è chiaro nel caso della luce monocromatica, composta da una radiazione con una lunghezza d'onda, o frequenza o energia, ben definita. Se la luce osservata è costituita da una miscela di radiazioni con lunghezza d'onda diversa il discorso si complica, ma sono sempre i contributi relativi delle diverse radiazioni che determinano il colore percepito.

Lunghezza d'onda, frequenza ed energia sono modi equivalenti per caratterizzare una radiazione. L'energia è direttamente proporzionale alla frequenza, mentre energia e frequenza sono inversamente proporzionali alla lunghezza d'onda. In altre parole, elevate frequenze e grandi energie corrispondono a piccole lunghezze d'onda, come nella parte violetta dello spettro visibile. Una unità di misura comune per le lunghezze d'onda è il nanometro (nm), che equivale a un miliardesimo di metro. Una unità di misura conveniente per l'energia della radiazione visibile è l'elettronvolt (eV), cioè l'energia acquistata da un elettrone accelerato dalla differenza di potenziale di un volt. In termini di lunghezza d'onda, lo spettro delle radiazioni visibili è compreso fra circa 700 nanometri, dove il rosso sfuma nell'infrarosso, e circa 400 nanometri, ai margini della radiazione ultravioletta. In termini di energia, tale spettro è compreso fra 1,77 e 3,1 elettronvolt.

In tutte le interazioni fra la radiazione elettromagnetica e la materia vale sempre la legge fondamentale della meccanica quantistica secondo la quale gli atomi possono esistere solo in stati discreti, ognuno caratterizzato da un ben definito livello di energia. I livelli intermedi sono proibiti. Il valore più basso possibile è detto stato fondamentale e gli altri livelli, detti stati eccitati, si dispongono sopra quello fondamentale formando come i gradini di una scala, il solo posto su cui si può appoggiare il piede. La distanza fra tali gradini è molto variabile. La luce o qualsiasi altra radiazione può venire assorbita solo se possiede esattamente l'energia necessaria per trasferire l'atomo da un livello a un altro superiore. Parimenti, quando un atomo decade da uno stato eccitato a uno più basso deve emettere una radiazione che porta con sé tutta l'energia che si deve perdere nel salto fra i due livelli. Questa energia si manifesta come fotone, o quanto di luce, la cui frequenza e lunghezza d'onda sono determinate dalla differenza di energia fra i due livelli.

Gli stati che presentano il maggiore interesse per l'origine del colore sono quelli che rappresentano i vari possibili livelli energetici degli elettroni. Negli atomi, ioni e molecole ogni elettrone deve occupare un orbitale che descrive una particolare distribuzione spaziale della carica dell'elettrone intorno al nucleo atomico. A loro volta gli orbitali sono organizzati in strati. Un ulteriore limite al numero degli stati possibili è posto dal fatto che ogni gradino della scala delle energie può venire occupato solo da un numero limitato di elettroni. In generale, procedendo dagli atomi più leggeri verso quelli più pesanti, gli elettroni si aggiungono in sequenza ordinata, a partire dai livelli più bassi. Il primo strato si riempie con due soli elettroni, i due immediatamente superiori richiedono ognuno otto elettroni. Negli strati pieni gli elettroni si dispongono a coppie e la configurazione dello strato è notevolmente stabile.

Per far saltare uno degli elettroni accoppiati di uno strato completo sul primo livello libero sovrastante occorre una quantità di energia non trascurabile. L'energia necessaria per una simile transizione può venire fornita, in genere, solo dalla radiazione ultravioletta o addirittura dalla zona X dello spettro. Ne segue che gli strati completi non contribuiscono direttamente alla formazione dei colori. Questa è dovuta generalmente alle transizioni degli elettroni spaiati, che molto spesso sono i più esterni. Si tratta in definitiva degli elettroni di valenza, cioè quelli responsabili della formazione dei legami chimici.

Le transizioni atomiche

Consideriamo un vapore dell'elemento sodio di densità sufficientemente bassa perché ogni atomo agisca indipendentemente da tutti i suoi vicini. L'atomo di sodio possiede 11 elettroni, ma 10 di essi fanno parte dei primi due strati completi. È quindi solo l'ultimo elettrone di valenza che prende parte alle interazioni dell'atomo con la luce. Quando l'atomo di sodio si trova nel suo stato fondamentale l'elettrone esterno occupa un orbitale che viene designato $3S_{1/2}$. I livelli energetici superiori più vicini (i primi scalini più in alto della scala) sono detti $3P_{1/2}$ e $3P_{3/2}$, cui corrispondono rispettivamente energie di 2,103 e 2,105 elettronvolt sopra quella dello stato fondamentale. Queste sono le più piccole quantità di energia che possano venire assorbite da un atomo di sodio nello stato fondamentale. Esse corrispondono a lunghezze d'onda di 589,6 e 589,1 nanometri, nella regione gialla dello spettro.

Al di sopra degli orbitali $3P$ si trova una moltitudine di altri stati eccitati corrispondenti a una maggiore distanza media degli elettroni dal nucleo e a una loro maggiore energia. Il numero di questi stati possibili è infinito, ma la distanza fra due stati successivi rimpicciolisce all'aumentare della energia, in modo che la serie converge a un limite finito. Per il sodio questo limite è pari a 5,12 elettronvolt e a tal punto l'elettrone più esterno non può più dirsi eccitato, ma è semplicemente del tutto isolato dall'atomo di appartenenza. In altre parole l'atomo è ionizzato.

Poniamo di avere un atomo di sodio ionizzato, per esempio da un quanto di radiazione ultravioletta, e che avvenga la ricombinazione fra lo ione e l'elettrone libero. Inizialmente l'elettrone può occupare uno degli orbitali più esterni, ma poi decade rapidamente allo stato fondamentale. Se questo decadimento avviene in una sola fase, dal livello di ionizzazione al livello fondamentale, l'atomo emetterebbe un solo fotone ultravioletto con un'energia di 5,12 elettronvolt. È però più probabile che il processo avvenga per stadi successivi, passando per livelli intermedi con salti accompagnati sempre dalla emissione di un fotone. Non tutti i possibili processi a cascata sono permessi. Delle «regole di selezione» determinano quelle consentite. La maggior parte delle sequenze permesse passano attraverso uno degli orbitali $3P$ e di qui, direttamente, allo stato fondamentale. Ne segue che fra i vari quanti di luce emessi si trovano quelli di energia pari a 2,103 e 2,105 elettronvolt, nello spettro del giallo. In effetti, queste due righe di emissione sono di gran lunga le più intense nello spettro di emissione degli atomi di sodio e un vapore di sodio eccitato o ionizzato emette una brillante luce gialla.

Questa caratteristica colorazione gialla della luce emessa dal sodio si osserva facilmente scaldando un sale di questo metallo in una fiamma a temperatura sufficientemente elevata per vaporizzarlo. In chimica analitica questa proprietà consente la determinazione sia qualitativa che quantitativa dei sali di sodio. Il doppietto di righe gialle predomina anche nello spettro delle lampade a vapori di sodio, nelle quali gli atomi del metallo vengono ionizzati da una scarica ad alta tensione.

Anche altri atomi producono delle righe di emissione caratteristiche quando vengono lasciati decadere allo stato fondamentale in seguito a eccitazione o ionizzazione. In ogni elemento, però, la distanza fra i livelli permessi è diversa e quindi è diverso il colore della luce emessa. Nel neon le righe più intense si trovano nella parte rossa dello spettro, il che spiega la colorazione rossa delle insegne luminose al neon. Il mercurio possiede le linee principali nelle regioni del verde e del viola dello spettro visibile e infatti le lampade a vapori di mercurio producono una luce bluastra carente di rosso e di giallo. I laser con cavità riempita di gas monoatomico utilizzano righe di emissione del tipo di quelle appena descritte. Anche i fulmini e gli archi elettrici emettono una luce dovuta a eccitazione degli atomi nel gas in cui vengono prodotti.

Lo spettro della luce solare ha un andamento molto vicino a quello della sensibilità dell'occhio umano; ne segue che la luce appare bianca quando la sua composizione spettrale è prossima a quella del Sole, mentre i colori possono venire definiti in base a come si scostano dallo spettro solare. La radiazione del Sole è circa uguale a quella di un corpo nero riscaldato alla temperatura di 5700 gradi centigradi. La forma dello spettro del corpo nero è determinata esclusivamente dalla temperatura. All'aumentare di questa essa diventa più stretta, con picco spostato verso le basse lunghezze d'onde. Quindi riscaldando un oggetto il suo colore cambia dal nero (nessuna emissione) al rosso, al giallo e infine all'azzurro pallido. Lo spettro può venire descritto come lunghezza d'onda, frequenza o energia, che sono solo mezzi alternativi per descrivere una radiazione.

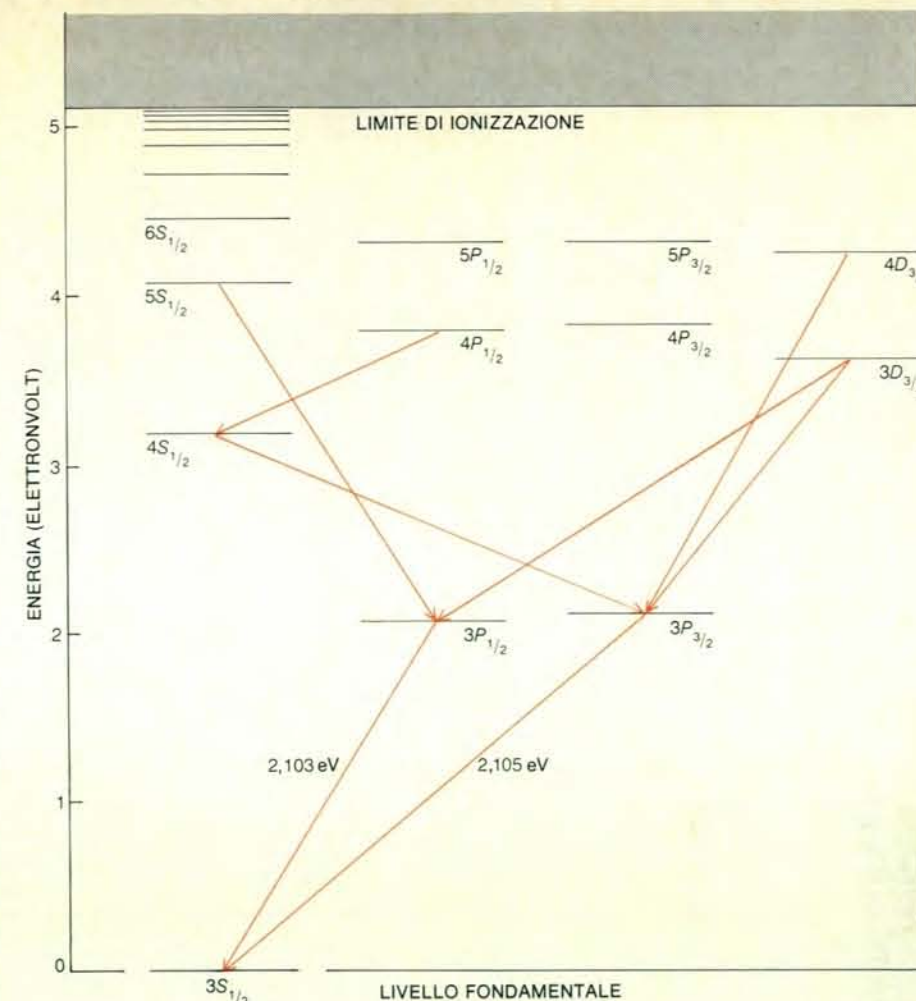
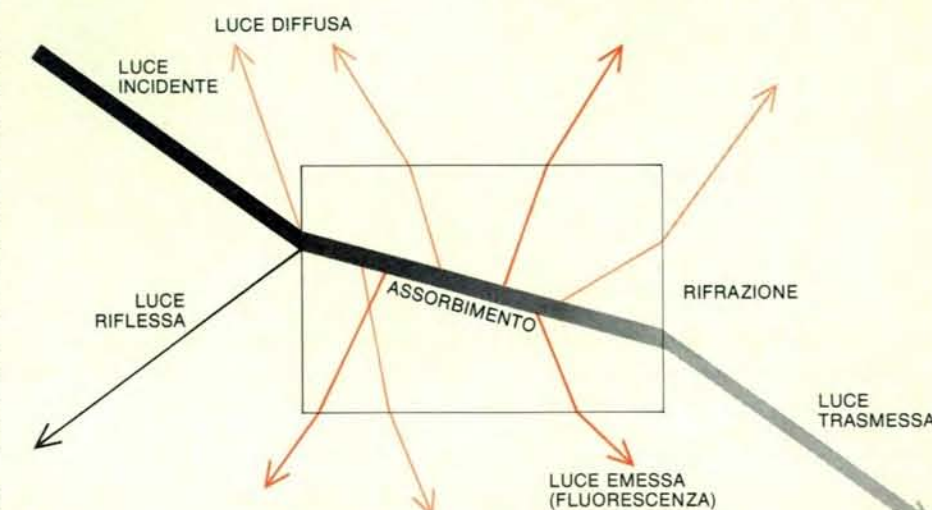
La radiazione del corpo nero

Le righe di emissione e di assorbimento con lunghezza d'onda ben definita sono tipiche dei gas. Lo spettro della luce emessa da un solido o un liquido è generalmente molto diversa, in quanto non presenta righe discrete ma è continuo in un intervallo di lunghezze d'onda.

La radiazione a carattere universale emessa dalla materia condensata è la radiazione del corpo nero, caratterizzata da uno spettro continuo con forma caratteristica. Per «corpo nero» intendiamo una sostanza ideale che assorbe o emette radiazioni di tutte le lunghezze d'onda senza privilegiarne alcuna. (I materiali reali possiedono tutti un'emissività inferiore a quella teorica ma molti approssimano lo spettro del corpo nero a temperatura elevata.) Questa radiazione ricopre un ruolo importante nella storia della fisica, dato che è stata l'analisi dello spettro del corpo nero che ha portato Max Planck a introdurre l'ipotesi dei quanti nel 1900. Egli trovò che solo ammettendo la quantizzazione dell'energia era possibile spiegare la forma dello spettro del corpo nero.

Nel corpo nero ideale lo spettro della radiazione non dipende dalla composizione chimica dell'emettitore e dipende da una sola variabile: la temperatura. Allo zero assoluto tutti gli atomi occupano il livello energetico più basso disponibile e non viene emessa radiazione. All'aumentare della temperatura alcuni atomi entrano in stati eccitati ma, dato che il processo di eccitazione è casuale, essi vengono a occupare un gran numero di stati, caratterizzati da un ampio spettro di energie. A qualsiasi temperatura finita il numero degli stati occupati aumenta gradualmente con l'energia fino a un massimo, poi decresce gradualmente. Quindi la forma dello spettro assomiglia a quello di un'onda in procinto di frangersi. La posizione della cresta e la forma più o meno ripida dell'onda dipendono dalla temperatura.

A temperatura ambiente le eccitazioni termiche sono confinate a livelli energetici bassi e viene emessa radiazione solo nella regione dell'infrarosso. Quando la temperatura raggiunge circa 700 gradi centigradi la maggior parte dell'emissione è ancora nell'infrarosso, ma comincia a comparire della radiazione visibile percepibile come una luminosità rossastra. All'aumentare della temperatura, il picco della curva si sposta verso energie più elevate e lunghezze d'onda inferiori, così che l'oggetto riscaldato diventa più luminoso e cambia colore. La serie dei colori va dal rosso all'arancio, al giallo, al bianco e infine all'azzurro chiaro, in accordo con le espressioni comuni «calor rosso», «calor bianco» e così via.

Nella fiamma di un pezzo di legno o di una candela le particelle incandescenti di carbone emettono una radiazione corrispondente a quella del corpo nero alla temperatura di non più di 1500 gradi centigradi, con una colorazione della luce che va dal rosso al giallo. Il filamento di tungsteno di una lampada a incandescenza ha

Il diagramma a scalini dei livelli di un atomo di sodio definisce uno spettro discreto di lunghezze d'onda, le sole che possono venire emesse o assorbite dall'atomo. Per salire a un livello più elevato l'atomo deve assorbire un quanto di radiazione, la cui energia corrisponde esattamente alla differenza fra le energie degli stati iniziale e finale. Ricadendo sul gradino inferiore l'atomo emette un quanto della stessa energia. La maggior parte delle transizioni in discesa passano attraverso i livelli designati $3P_{1/2}$ e $3P_{3/2}$ fino allo stato fondamentale $3S_{1/2}$. In queste transizioni vengono emessi dei quanti con energie di 2,103 e 2,105 elettronvolt, nella regione gialla dello spettro, e infatti i vapori di sodio emettono una luce gialla intensa. Nel diagramma a scalini hanno significato solo le distanze in verticale, la separazione orizzontale dei livelli è fatta solamente per chiarezza.

Le interazioni della luce con la materia comprendono la riflessione, la rifrazione, la diffusione e l'assorbimento. Parte della radiazione assorbita può anche venire riemessa come fluorescenza. Gli effetti di queste interazioni possono dipendere dalla lunghezza d'onda della radiazione che interagisce, dando origine a dei colori. L'assorbimento selettivo di lunghezze d'onda brevi può rendere un oggetto giallo, arancio o rosso, a seconda dell'estensione dell'assorbimento. Gli stati condensati della materia assorbono le bande di lunghezza d'onda ampie e continue ma non righe discrete.

Il campo elettrico entro un cristallo può influire sul colore alterando lo stato degli atomi o ioni entro il reticolo. I colori del rubino e dello smeraldo sono dovuti all'assorbimento di determinate lunghezze d'onda da parte degli elettroni spaiati negli ioni cromo. Le transizioni che provocano l'assorbimento sono le stesse nei due casi: gli ioni passano dallo stato fondamentale $4A_2$ ai livelli eccitati $4T_2$ e $4T_1$. Le energie di questi stati sono stabilite dall'intensità del campo cristallino. Nel rubino le bande di assorbimento bloccano la luce viola, verde e gialla. Vengono trasmessi invece il rosso e un po' di blu che danno alla gemma il colore caratteristico. Il campo cristallino dello smeraldo è più debole e ciò abbassa entrambe le bande di assorbimento. Vengono così eliminata la trasmissione del rosso e migliorata quella del blu, che si estende nel verde. In entrambi i casi uno ione eccitato torna allo stato fondamentale attraverso il livello intermedio $2E$, la cui energia è poco influenzata dal campo cristallino. Le transizioni dal livello $2E$ a quello fondamentale danno luogo alla fluorescenza rossa quasi identica delle due pietre.

una temperatura di circa 2200 gradi centigradi e fornisce una luce calda bianco-giallastra. Una lampada flash, che può raggiungere una temperatura di 4000 gradi, produce una luce ancora più bianca.

Lo spettro della luce solare è molto simile alla curva del corpo nero. La sua forma è determinata dalla temperatura esistente sulla superficie del Sole, che è di circa 5700 gradi centigradi. Questo spettro possiede un picco allargato centrato sui 2,2 elettronvolt, pari a 560 nanometri, nella zona del giallo-verde. Il massimo di sensibilità dell'occhio umano cade appunto in quella zona. In effetti l'idea di bianco pare dipendere in gran parte dallo spettro della luce del giorno, che è dominata dalla radiazione del Sole. Grosso modo la luce viene percepita come bianca se il suo spettro somiglia a quello della luce solare. Gli altri colori possono venire definiti in base allo scostamento dallo spettro solare.

Il campo cristallino

Quando gli atomi si combinano a formare una molecola o condensano formando un liquido o un solido vengono introdotte nuove possibilità di eccitazione. Fra queste vi sono vibrazioni e rotazioni meccaniche che sono impossibili in un atomo isolato. Per esempio, gli atomi di una molecola biatomica possono oscillare come se fossero connessi con una molla e possono ruotare intorno al loro centro di massa comune. Tali movimenti possono avere influenza sul colore di una sostanza. Nell'acqua, per esempio, una distorsione complessa delle molecole comporta l'assorbimento di un po' di energia all'estremità rossa dello spettro ed è quindi la causa del colore azzurrino dell'acqua e del ghiaccio. Per la maggior parte, però, l'energia delle eccitazioni vibrazionali e rotazionali è piccola e viene dissipata sotto forma di radiazione infrarossa, o calore.

Un'altra conseguenza dei legami fra atomi è il cambiamento dello stato degli elettroni di valenza. Negli atomi isolati questi sono spaiati e costituiscono la causa principale dell'origine dei colori. In una molecola e in molti solidi, invece, gli elettroni di valenza di un atomo si accoppiano con quelli degli atomi adiacenti. Sono queste coppie che producono il legame chimico. In conseguenza di tali accoppiamenti, le bande di assorbimento degli elettroni di valenza vengono spostate nella regione dell'ultravioletto e non contribuiscono più alla formazione dei colori. Solo alcuni elettroni in stati eccezionali rimangono disponibili a questo fine. È, comunque, evidente che tali stati eccezionali non sono poi troppo rari. Se così fosse, la maggior parte delle sostanze sarebbero trasparenti per la radiazione visibile.

Un insieme di stati elettronici poco comuni si trova nei metalli di transizione, quali il ferro, il cromo, il rame e gli elementi delle terre rare. Gli atomi degli elementi di transizione possiedono degli strati interni che rimangono parzialmente vuoti. Essi possiedono quindi degli elettroni spaiati i cui stati eccitati cadono spesso nella zona visibile dello spettro, dando origine a colori molto intensi. Per esempio, il rubino sia lo smeraldo devono i loro colori alla presenza di tracce di cromo.

Il componente base del rubino è il corindone, un ossido di alluminio dalla formula Al_2O_3. Il corindone puro è incolore, ma nel rubino la sostituzione di una piccola percentuale degli ioni alluminio con ioni cromo (Cr^{+++}) produce il colore caratteristico della gemma. Gli ioni cromo possiedono tre elettroni spaiati il cui minimo livello energetico corrisponde a uno stato fondamentale designato $4A_2$, al di sopra del quale si trova uno spettro complicato di stati eccitati. Tutti gli stati eccitati non corrispondono a un'energia ben definita ma si allargano formando delle bande, la cui struttura è a sua volta modificata dalla presenza degli altri ioni nel reticolo cristallino. Per quanto l'identità dei vari stati sia definita dalla configurazione elettronica dello ione cromo, la posizione assoluta di ogni livello nello spettro energetico è determinata dal campo elettrico in cui è immerso lo ione. L'intensità e la simmetria di questo sono a loro volta determinati dalla natura e dalla disposizione spaziale degli ioni che circondano quello di cromo.

Nel rubino gli ioni cromo si trovano al centro di un ottaedro distorto, costituito da sei atomi d'ossigeno. I legami interatomici sono lunghi circa 0,19 nanometri e possiedono circa il 63 per cento di carattere ionico. Questo vuol dire che i doppietti elettronici che costituiscono il legame passano più tempo nei pressi degli ioni ossigeno che vicino agli ioni alluminio o cromo. Questa distribuzione di cariche elettriche origina un campo elettrico relativamente intenso, chiamato campo cristallino o campo del legante. Immergendo uno ione cromo in tale campo, tre stati eccitati dei suoi elettroni spaiati possiedono energie nel campo del visibile.

I tre stati eccitati vengono designati $2E$, $4T_2$ e $4T_1$. Le regole di selezione proibiscono una transizione diretta dallo stato fondamentale al livello $2E$, ma entrambi gli stati $4T$ possono venire raggiunti da quello fondamentale. Le energie associate a queste transizioni corrispondono a lunghezze d'onda delle regioni violetta e giallo-verde dello spettro. Dato che i livelli citati non sono delle righe sottili, bensì delle bande molto larghe, può venire assorbito uno spettro di radiazioni relativamente ampio. Quindi, quando la luce bianca attraversa un cristallo di rubino, ne emerge impoverita di radiazioni

In un cristallo si può formare un centro di colore quando un elettrone prende il posto di uno ione dislocato. Nel caso tipico della fluorite, fluoruro di calcio (CaF_2), l'elettrone sostituisce uno ione fluoro. Questo elettrone possiede uno spettro di stati eccitati che si estende nel campo delle energie del visibile. I centri di colore presenti nella fluorite, detti centri *F*, conferiscono al minerale un colore porporino. La struttura originaria del cristallo può venire ristabilita tramite riscaldamento, con la conseguente scomparsa del colore. Esistono anche centri di colore formati da elettroni o da «buche», derivati dallo spostamento di un solo elettrone anziché di un intero ione.

Gli orbitali molecolari descrivono il comportamento di doppietti elettronici distribuiti fra diversi atomi o un'intera molecola. Essi presentano spesso dei forti assorbimenti nel visibile. Il doppietto che forma un legame chimico assorbe normalmente solo nella banda dell'ultravioletto, ma quando esistono configurazioni di legami equivalenti che coinvolgono più atomi l'energia di eccitazione del doppietto viene ridotta. I più frequenti orbitali molecolari che provocano la formazione di colori sono quelli associati a sistemi di legami coniugati (legami singoli e doppi alternati) presenti in composti organici quale il benzene *(a)*. Spostando tre coppie di elettroni la configurazione di legami del benzene viene capovolta. Una rappresentazione migliore della struttura di questa molecola *(b)* mostra dei legami singoli fra tutti gli atomi di carbonio, con i tre doppietti in più distribuiti in tutta la molecola in orbitali «π». Questi orbitali sono i responsabili della produzione dei colori. Nel benzene i livelli eccitati degli orbitali π si trovano ancora nell'ultravioletto, e infatti il benzene è incolore, ma in molecole contenenti sistemi coniugati più estesi, come nel colorante violetto in cristalli *(c)*, le bande di assorbimento cadono nel visibile. Il colore può venire intensificato da gruppi chimici, detti auxocromi, che accettano o donano elettroni.

corrispondenti al viola e al giallo-verde. Invece viene trasmesso praticamente tutto il rosso e un po' di blu, il che origina il colore rosso intenso con una sfumatura porpora del rubino.

Le regole di selezione impongono agli elettroni di rientrare dai livelli eccitati $4T$ allo stato fondamentale $4A_2$ attraversando il livello intermedio $2E$. Le transizioni iniziali dai livelli $4T$ al $2E$ liberano quantità molto piccole di energia, corrispondenti all'infrarosso, ma il decadimento dallo stato $2E$ allo stato fondamentale produce una forte emissione di luce rossa. Va notato che questa radiazione, a differenza delle bande trasmesse, non è presente nel fascio incidente sul cristallo. Viene prodotto all'interno di questo per fluorescenza. In effetti la fluorescenza rossa del rubino si osserva in maniera particolarmente chiara illuminandolo con luce verde o violetta o con radiazione ultravioletta. D'altro canto, questa fluorescenza può venire spenta da impurezze di ferro, spesso presenti nei rubini naturali. La radiazione di un laser a rubino deriva dalla fluorescenza rossa di cristalli sintetici, del tutto privi di ferro.

Rubino e smeraldo

La varietà dei colori prodotti dal campo cristallino risulta evidente confrontando il rubino con lo smeraldo. L'impurezza che produce il colore dello smeraldo è ancora lo ione Cr^{+++} che sostituisce piccole quantità dello ione alluminio (Al^{+++}). Le somiglianze fra le due sostanze sono ancora più strette: anche nello smeraldo gli ioni cromo sono al centro di sei ioni ossigeno in configurazione ottaedrica e la lunghezza dei legami è sempre di circa 0,19 nanometri. Nello smeraldo, però, il reticolo cristallino fondamentale è quello di un alluminosilicato di berillio, $Be_3Al_2Si_6O_{18}$, e la differenza più significativa è costituita dalla natura dei legami chimici, che sono meno ionici di qualche unità percentuale, cosa che diminuisce l'intensità del campo elettrico in cui sono immersi gli ioni cromo. Ne risulta che i due livelli $4T$ vengono a trovarsi a livelli energetici un po' più bassi, mentre la posizione della banda $2E$ è praticamente inalterata. L'effetto principale di questi mutamenti è di spostare la banda di assorbimento che nel rubino blocca il giallo e il verde verso le energie più basse della zona gialla e rossa dello spettro. Lo smeraldo, quindi, assorbe per lo più luce rossa e trasmette molto più del rubino quella blu e verde.

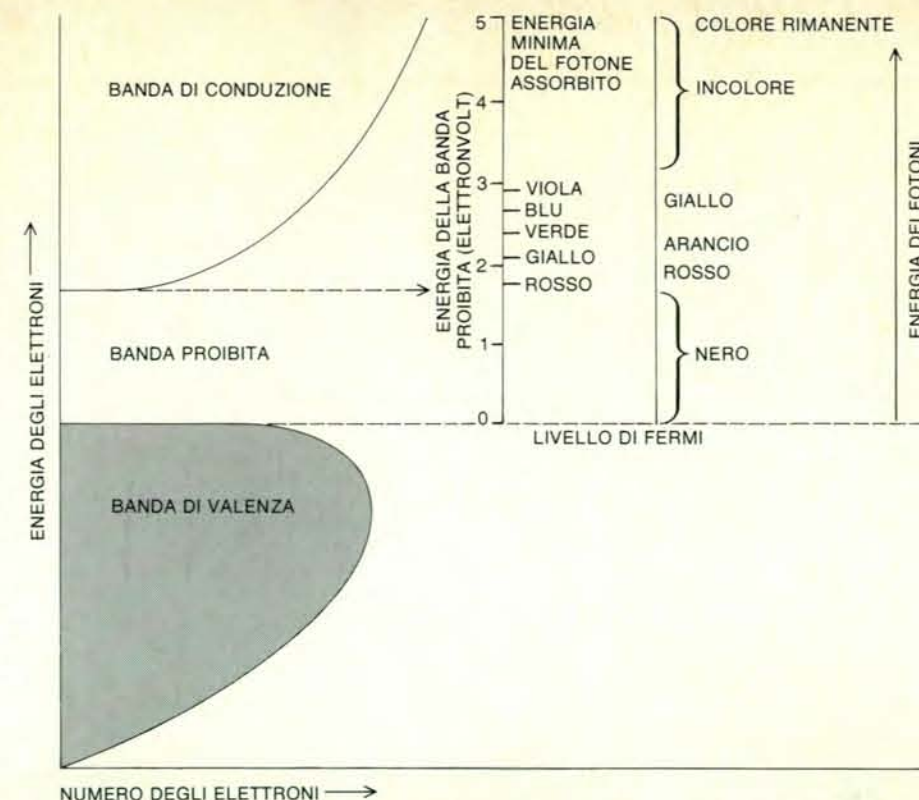
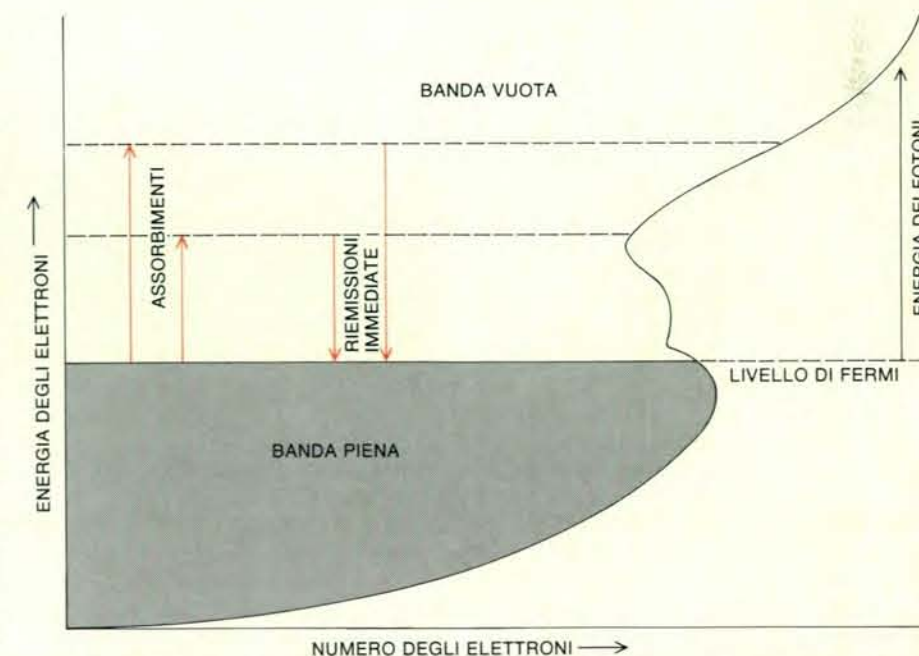
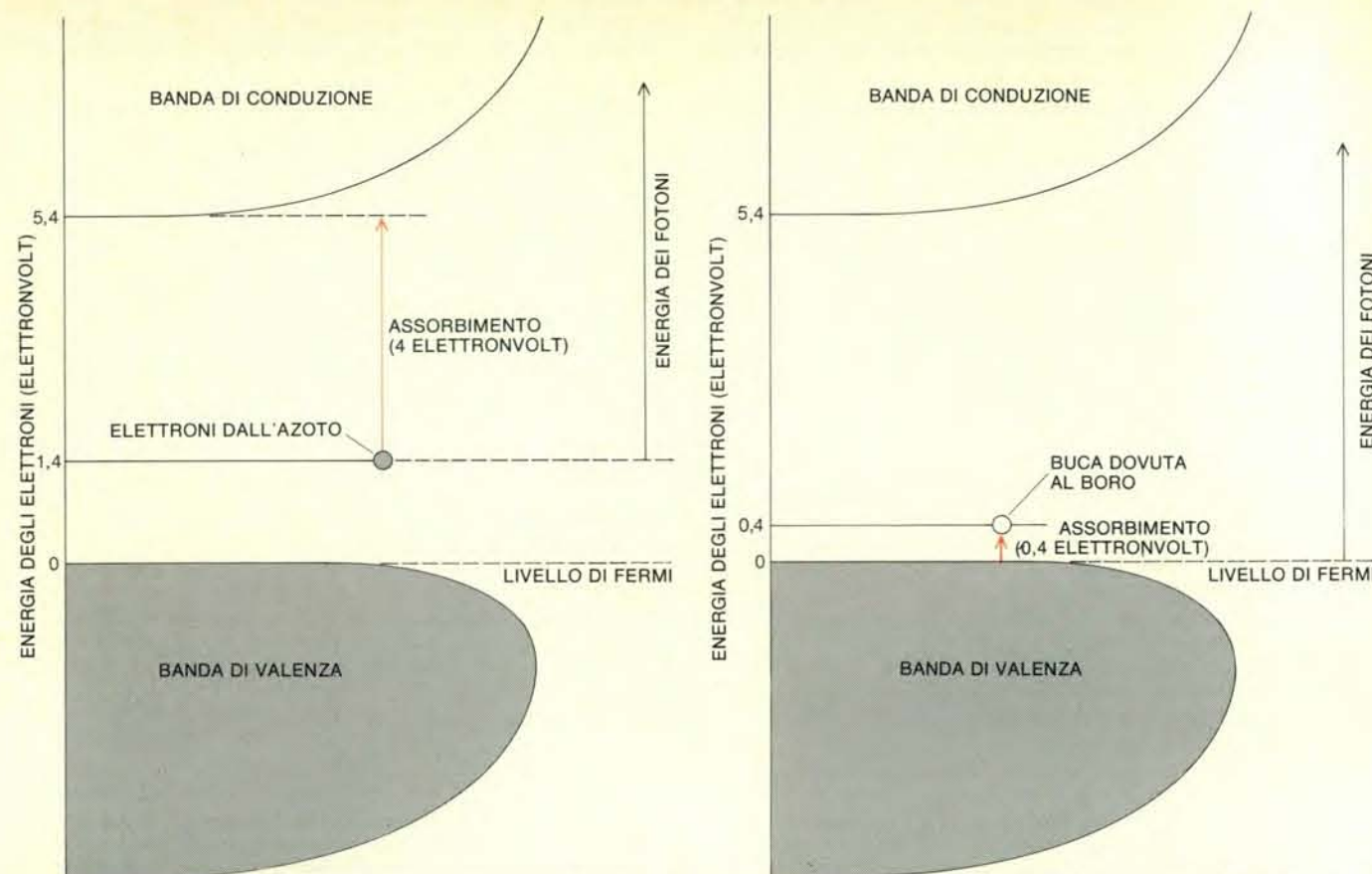
Singolarmente la fluorescenza dello smeraldo è quasi identica a quella del rubino. Il motivo sta nel fatto che l'energia del livello $2E$ viene influenzata di poco dalla riduzione del campo cristallino. La somiglianza delle emissioni di fluorescenza dimostra che il fatto che il rubino abbia colore e fluorescenza entrambi rossi è una pura coincidenza.

In una posizione di composizione spettrale intermedia, fra il rubino e lo smeraldo, si trova la rara (e perciò preziosa) pietra alessandrite. Ancora una volta il colore è dovuto alla presenza di ioni cromo che sostituiscono l'alluminio, ma in questo caso la matrice cristallina è quella di un alluminato di berillio, $BeAl_2O_4$. Il campo cristallino che stabilisce i livelli energetici degli ioni cromo è più forte che nello smeraldo, ma più debole che nel rubino, con il risultato che le bande di trasmissione nel verde e nel rosso sono ben bilanciate. La quasi uguaglianza di tali bande ha una conseguenza straordinaria: alla luce del Sole, ricca di blu, la gemma appare di colore verde-azzurro; alla luce più rossa di una candela o di una lampada a incandescenza appare rossa.

Colori dovuti al campo cristallino possono aversi tutte le volte che in un solido vengono a trovarsi degli ioni con elettroni spaiati. L'acquamarina, la giada e il quarzo citrino devono i loro colori a un meccanismo simile a quello visto per il rubino e lo smeraldo, ma questa volta l'impurezza di metallo di transizione è costituita da ioni ferro, anziché cromo. Molti composti e cristalli nei quali i metalli di transizione sono un componente fondamentale, e non solo una impurezza, possiedono pure una netta colorazione. Fra i minerali che cadono entro questa categoria troviamo l'azzurrite blu o verde, il turchese e la malachite, il cui colore è dovuto al rame e i granati rossi, nei quali il colore è prodotto dal ferro. La maggior parte dei pigmenti usati nelle pitture sono composti di metalli di transizione.

Centri di colore

L'assorbimento degli elettroni nei metalli di transizione non è il solo meccanismo di produzione dei colori in presenza del campo cristallino. In effetti, gli elettroni non debbono per forza appartenere a un determinato atomo. Un elettrone in eccesso, non legato ad alcun atomo, può essere sufficiente, se si trova intrappolato in qualche difetto strutturale del reticolo, per esempio in corrispondenza di uno ione mancante o di un'impurezza. Lo stesso effetto può avere anche una «buca», cioè la mancanza di un elettrone di un doppietto. Anomalie di questo tipo sono dette centri di colore o centri F (dal tedesco *Farbe*, colore).

Sono noti molti tipi di centri di colore, ma solo in pochi casi si è compreso il meccanismo di produzione del colore. Uno di questi è il centro F porpora della fluorite, un minerale nel quale esistono molti centri di colore diversi. La fluorite è fluoruro di calcio, CaF_2, in cui normalmente ogni ione calcio è circondato da otto ioni fluoro. Un centro F si forma quando viene a mancare uno ione fluoro nel reticolo. Questo perdita può verificarsi in vari modi diversi: quando il cristallo cresce in un eccesso di ioni calcio; esponendolo a radiazione di energia elevata (che può spostare uno ione dalla sua posizione normale); o applicando un campo elettrico molto intenso (che sposta il fluoro per elettrolisi). Per conservare la neutralità elettrica del cristallo, il posto lasciato vacante nel reticolo dallo ione fluoro deve venire occupato da qualche altra

La struttura elettronica di un metallo è caratterizzata da una banda essenzialmente continua di livelli energetici permessi. La banda è riempita dallo stato fondamentale fino a un valore di energia detto livello di Fermi. Tutti gli stati più elevati sono vuoti e quindi possono accettare elettroni eccitati. La configurazione permette di assorbire tutte le radiazioni, dal lontano infrarosso fino all'ultravioletto e oltre. Un materiale che assorbe tutto il visibile dovrebbe essere nero. Ciò non avviene perché un elettrone eccitato può tornare subito allo stato di partenza riemettendo un quanto con la stessa lunghezza d'onda di quello assorbito. Le superfici metalliche sono quindi molto riflettenti.

La struttura a bande di un semiconduttore è simile a quella di un metallo, con la differenza che esiste una banda di livelli proibiti che separa la banda di valenza e quella di conduzione. Per poter venire assorbito, un quanto di radiazione deve possedere una energia non inferiore a quella necessaria per far saltare un elettrone dal livello più elevato della banda di valenza a quello più basso della banda di conduzione. Il colore di un semiconduttore puro è determinato dall'ampiezza della banda proibita. Se questa corrisponde a energie nell'infrarosso, tutto il visibile viene assorbito e il materiale appare nero. Se l'energia di salto cade nel visibile vengono assorbite solo le lunghezze d'onda inferiori e il semiconduttore può avere un colore variabile dal giallo al rosso. Se infine l'energia di salto è nell'ultravioletto, tutto il visibile viene trasmesso e il materiale appare incolore.

Il drogaggio di un semiconduttore con opportune impurezze in tracce provoca la formazione di livelli permessi all'interno della banda proibita. Nel diamante, che è un semiconduttore con ampiezza della banda proibita pari a 5,4 elettronvolt, il drogaggio con azoto (in sostituzione di pochi atomi di carbonio) introduce un livello di stati riempiti a 1,4 elettronvolt sopra il livello di Fermi. Un elettrone di tali stati può venire spedito nella banda di conduzione in seguito all'assorbimento di un quanto di radiazione di 4 elettronvolt. Dato che il livello in questione ha una certa ampiezza, si verifica un assorbimento nel viola e il diamante assume una colorazione gialla. L'aggiunta di boro al diamante produce delle buche, o stati vuoti, a un livello centrato a 0,4 elettronvolt sopra il livello di Fermi. Un diamante con questa impurezza è blu perché un elettrone della banda di valenza può saltare sul livello in questione assorbendo un quanto di luce di grande lunghezza d'onda.

carica negativa. Quando questa è costituita da un elettrone libero si forma un centro *F*. L'elettrone viene tenuto fermo non perché appartiene a un nucleo centrale, come in un atomo o uno ione, bensì dal campo cristallino creato da tutti gli ioni circostanti. In questa situazione esso può occupare uno stato fondamentale oppure vari stati eccitati, simili a quelli visti per i metalli di transizione. Lo spostamento degli elettroni fra tali stati produce sia il colore sia la fluorescenza.

Il colore del quarzo affumicato (morione) è attribuito a un centro di colore costituito da una buca. Il reticolo del quarzo è costituito da biossido di silicio (SiO_2), ma un prerequisito essenziale per la formazione di tale centro di colore è la presenza di impurezze di alluminio che sostituiscono qualche atomo di silicio. Dato che l'alluminio ha valenza +3, mentre per il silicio la valenza è +4, per mantenere la neutralità elettrica occorre che sia presente nelle vicinanze uno ione di un metallo alcalino o anche un idrogenione. Il quarzo naturale è quasi sempre contaminato con impurezze di alluminio, ma questo da solo non è in grado di fargli assumere una colorazione perché non possiede elettroni spaiati. Il centro di colore si forma in seguito all'esposizione per alcuni minuti a una sorgente intensa di raggi X o gamma. Ma può bastare una sorgente molto debole se l'esposizione ha durate di tipo geologico. La radiazione espelle un elettrone da un doppietto in un atomo di ossigeno vicino a un'impurezza di alluminio, lasciando nell'orbitale un elettrone spaiato. L'elettrone mancante viene detto buca e quello che rimane assume la possibilità di spostarsi lungo una serie di livelli eccitati, simili a quelli dell'elettrone in eccesso dei centri *F*.

L'ametista è un'altra varietà di quarzo che deve il suo colore a un centro formato da una buca, ma l'impurezza, in questo caso, è costituita da ferro anziché alluminio. Certe vecchie bottiglie fatte con un vetro contenente ferro e manganese diventavano di color porpora dopo molti anni di esposizione alla luce solare intensa, a causa della formazione di centri di colore. È il cosiddetto vetro ametista del deserto. Lo stesso effetto si ottiene esponendo per 10 minuti quella qualità di vetro ai raggi gamma emessi dal cobalto 60.

La maggior parte dei centri di colore è stabile a temperature non troppo elevate. Nella fluorite l'innalzamento della temperatura mobilizza gli ioni fluoro spostati, consentendo loro di riguadagnare la posizione iniziale nel reticolo. L'ametista, riscaldata, cambia colore diventando o quarzo citrino (quarzo topazio) o un raro quarzo verde. Entrambi questi colori sono dovuti all'influenza del ferro senza la presenza dei centri di colore ametista. In alcuni casi anche la luce solare può produrre la scomparsa dei centri di colore. Per esempio, un topazio incolore può venire irraggiato con formazione di un colore bruno-arancione intenso, ma dopo pochi giorni di irraggiamento la colorazione scompare. Il topazio naturale del medesimo colore è assolutamente stabile.

Orbitali molecolari

È già stato detto che nelle molecole e nei solidi gli elettroni di valenza impegnati nei legami chimici formano dei doppietti stabili i cui stati eccitati si trovano nello ultravioletto. In realtà questo è vero solo quando il doppietto di elettroni rimane confinato in un determinato legame fra due atomi. In molti casi gli elettroni possono spostarsi a distanze molto grandi entro l'intera molecola o addirittura entro un solido macroscopico. In questi casi non sono legati strettamente ai loro atomi di appartenenza e l'energia necessaria per saltare in uno stato eccitato è molto ridotta. Si dice allora che gli elettroni occupano degli orbitali molecolari, anziché atomici, e danno origine a moltissimi colori.

Un meccanismo con il quale gli orbitali molecolari possono produrre dei colori è il trasferimento di carica da uno ione all'altro. Un esempio di questo processo è lo zaffiro blu. Come nel rubino, la matrice dello zaffiro è il corindone, questa volta con due impurezze significative, ferro e titanio, entrambe in posizioni normalmente occupate dall'alluminio. Nello stato energetico più basso il ferro ha una carica +2 e il titanio +4. Si ha uno stato eccitato quando un elettrone viene trasferito dal ferro al titanio, così che entrambi gli ioni assumono una carica +3. Per questo trasferimento di carica occorre un'energia di circa 2 elettronvolt. La transizione provoca la formazione di un'ampia banda di assorbimento che si estende dal giallo al rosso, producendo il colore blu intenso dello zaffiro.

In alcune sostanze il ferro è presente in entrambe le sue valenze comuni, Fe^{++} e Fe^{+++}. Trasferimenti di carica fra questi ioni danno origine a colori che vanno dal blu scuro al nero, come nel minerale di ferro nero magnetite.

La teoria degli orbitali molecolari si applica anche ai colori osservati in molte sostanze organiche in cui gli atomi di carbonio (e talora di azoto) sono legati mediante un sistema di legami alternativamente singoli e doppi, detti legami coniugati. L'esempio più conosciuto di sistema coniugato è l'anello del benzene, con sei atomi di carbonio, ma ne esistono molti altri. Dato che ogni legame è dovuto a un doppietto elettronico posto in comune fra due atomi, lo spostamento di un doppietto da un doppio legame al legame singolo adiacente provoca il capovolgimento dell'intera sequenza di legami. Le due strutture così definite sono equivalenti e non vi è ragione per sceglierne una fra le due. In effetti, la migliore rappresentazione di tale struttura mostra tutti gli atomi connessi con un solo legame, mentre i restanti doppietti sono distribuiti uniformemente nel sistema coniugato in orbitali molecolari che, in questo caso, vengono denominati orbitali π.

La natura distribuita degli orbitali π in un sistema di legami coniugati tende a far diminuire l'energia di eccitazione dei doppietti di elettroni. Nel benzene l'energia del più basso livello eccitato cade ancora nell'ultravioletto, e infatti il benzene è incolore, ma in molecole più grandi, e in particolare in quelle costituite da più anelli, l'assorbimento può estendersi nel visibile. Queste molecole organiche colorate vengono dette cromofore. Il colore può venire esaltato legando al sistema coniugato dei gruppi laterali, detti auxocromi, in grado di donare o di accettare elettroni.

Numerosi pigmenti di origine biologica devono il loro colore a estesi sistemi di orbitali π. Fra questi la clorofilla verde delle piante e l'emoglobina rossa del sangue. I coloranti organici si basano sul medesimo meccanismo. Il *Color Index* della Society of Dyers and Colourists elenca circa 8000 sostanze di questo tipo.

Talune specie chimiche con orbitali molecolari estesi sono in grado di emettere luce di fluorescenza. Come nel rubino e nello smeraldo, la fluorescenza ha luogo quando uno stato eccitato decade a quello fondamentale passando attraverso uno stato intermedio e almeno una delle transizioni corrisponde a una lunghezza d'onda dello spettro visibile. Gli sbiancanti ottici aggiunti ai detersivi raggiungono il loro effetto assorbendo l'ultravioletto presente nella luce del Sole e riemettendo parte dell'energia assorbita sotto forma di radiazione nella zona blu dello spettro. I laser a colorante, nei quali l'elemento attivo è un pigmento organico, operano secondo il medesimo principio.

La fluorescenza si osserva anche in cer-

La figura mostra i colori della banda proibita in sei semiconduttori. Il solfuro di cadmio puro *(in basso a sinistra)* ha una banda proibita che si estende per 2,6 elettronvolt e può quindi assorbire solo il viola. Questa sostanza ha quindi un colore giallo complementare al viola. Nel seleniuro di cadmio puro *(in basso a destra)*, l'estensione della banda proibita è di 1,6 elettronvolt. Tutto il visibile può venire assorbito e il colore è quindi nero. Gli altri quattro materiali sono miscele variabili dei due composti di cadmio che mostrano un diverso assorbimento, con un colore che si sposta dall'arancio al rosso. Vengono usati come pigmenti con il nome di giallo cadmio e rosso cadmio.

ti sistemi di orbitali molecolari in cui l'energia richiesta per produrre lo stato eccitato iniziale non proviene da una radiazione. Per esempio, la bioluminescenza delle lucciole e di certi pesci abissali è prodotta da una serie di reazioni chimiche che provocano l'eccitazione di una molecola con orbitali π estesi. Le lampade a luce fredda da usare e gettare imitano il processo biologico.

Metalli e semiconduttori

L'estensione spaziale degli orbitali elettronici raggiunge il culmine nei metalli e nei semiconduttori. In queste sostanze gli elettroni sono completamente liberi da qualsiasi legame con un atomo particolare e possono muoversi liberamente entro volumi macroscopici. La loro possibilità di spostamento è limitata solo dalle dimensioni del materiale in cui si trovano. Il numero enorme di elettroni mobili (dell'ordine di 10^{23} per centimetro cubo) è la causa delle eccezionali proprietà ottiche ed elettriche dei metalli e dei semiconduttori.

In un metallo tutti gli elettroni di valenza sono essenzialmente equivalenti, visto che possono liberamente scambiarsi di posto. Si potrebbe quindi supporre che tutti possiedono la medesima energia, ma una regola della meccanica quantistica lo proibisce. Come in un atomo isolato, anche nei solidi solo un numero limitato di elettroni può occupare il medesimo livello energetico. Ne discende che in un metallo devono esistere molti livelli energetici estremamente ravvicinati. In effetti essi formano un continuo a partire dal livello fondamentale.

Alla zero assoluto questa banda continua di stati è riempita dal livello più basso fino a un valore di energia detto livello di Fermi. Tutti gli stati al di sopra di questo sono vacanti. Qualunque immissione di energia, non importa quanto piccola, spinge un elettrone a occupare uno dei livelli liberi sopra il livello di Fermi.

Dato che i metalli possiedono un continuo di stati eccitati, essi sono in grado di assorbire radiazioni di qualsiasi lunghezza d'onda. Una superficie dotata della proprietà di assorbire tutta la radiazione visibile dovrebbe apparire nera, ma ciò non si verifica per i metalli. La ragione va cercata ancora nella mobilità degli elettroni dei metalli. Quando un elettrone assorbe un fotone e salta su uno stato eccitato può immediatamente riemettere un fotone della medesima energia e tornare al livello di partenza. A causa di tale rapido ed efficiente reirraggiamento, la superficie di un metallo pare quindi più riflettente che assorbente e in effetti queste sostanze possiedono un caratteristico aspetto lucido. Se la superficie metallica è liscia, la riflessione può essere speculare.

La diversità nei colori dei metalli, come per esempio fra oro e argento, è dovuta a differenze nel numero degli stati disponibili a particolari energie sopra il livello di Fermi. Dato che la densità degli stati non è uniforme, alcune lunghezze d'onda vengono assorbite e riemesse con maggiore efficienza di altre. Oltre ad assorbirla e riemetterla, certi metalli possono anche trasmettere la luce. La trasmissione è però sempre molto debole e può venire osservata solo in strati sottili. Essa dipende dalla lunghezza d'onda e quindi origina dei colori. Una lamina di oro battuto, nella quale la struttura cristallina è molto distorta, trasmette solo la luce verde. L'oro colloidale nel vetro rubino non è sottoposto a tensioni interne e trasmette luce rosso-porpora.

I colori della banda proibita

Bande ampie e continue di stati elettronici sono caratteristiche anche di un'altra classe di materiali: i semiconduttori. La proprietà distintiva di questi materiali è che il numero medio di elettroni di legame per ogni atomo è esattamente quattro. Fra essi troviamo forme cristalline di alcuni elementi del quarto gruppo del sistema periodico, quali il silicio, il germanio e la fase diamante del carbonio. Anche vari composti, come l'arseniuro di gallio, sono dei semiconduttori.

I semiconduttori si differenziano dai metalli perché la banda di livelli consentiti è divisa in due parti. Tutti i livelli energetici inferiori formano una banda di valenza, che allo stato fondamentale è completamente piena. Tutti gli stati eccitati si trovano in una banda di conduzione separata, che nello stato fondamentale è completamente vuota. Le due bande sono separate da un intervallo di livelli non accessibili, detto banda proibita. Questa suddivisione in bande ha effetti importanti sulle proprietà ottiche dei semiconduttori. Gli elettroni non possono più assorbire quanti di energia arbitrariamente bassa. L'energia minima assorbibile è quella necessaria per far saltare un elettrone dal livello superiore della banda di valenza a quello inferiore della banda di conduzione.

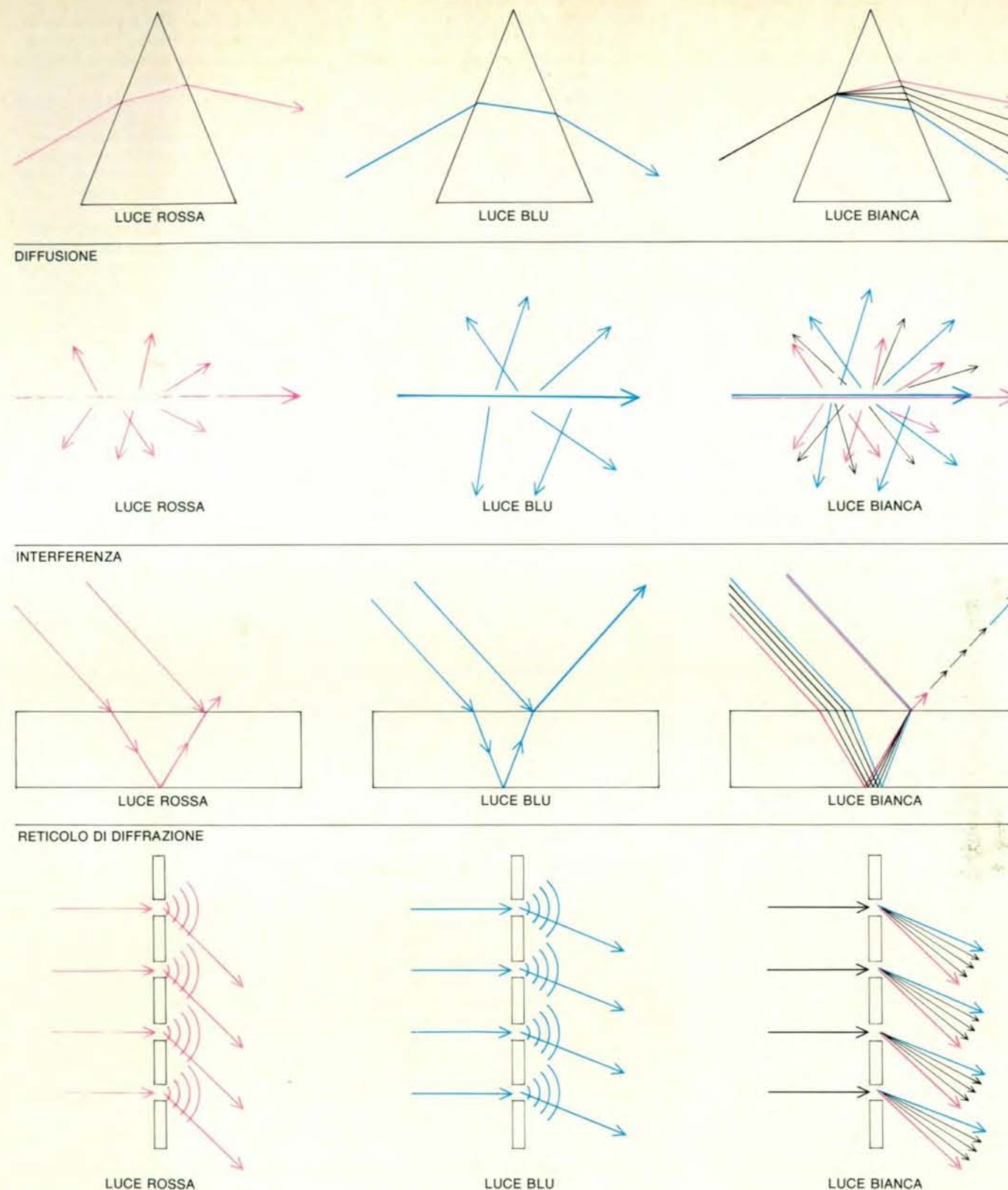
Il colore di un semiconduttore puro dipende solo dall'ampiezza della banda proibita. Se questo intervallo di energia è più piccolo della minore energia presente nello spettro visibile, allora tutte le lunghezze d'onda nel visibile vengono assorbite. Semiconduttori con piccola banda proibita nei quali il processo di riemissione è rapido ed efficace, come il silicio, possiedono lucentezza metallica ma altri semiconduttori di questo tipo sono neri. All'estremo opposto la banda proibita può essere maggiore della massima energia presente nella radiazione visibile: in questo caso la radiazione visibile non viene assorbita e il semiconduttore è privo di colore. È il caso del diamante che, con un'ampiezza della banda proibita di 5,4 elettronvolt, è trasparente alla luce visibile e, parzialmente, anche all'ultravioletto.

Quando l'energia della banda proibita cade entro lo spettro visibile il semiconduttore ha un colore definito. Il cinabro, un minerale di mercurio con formula HgS, noto anche come vermiglione, possiede una banda proibita dell'ampiezza di 2,1 elettronvolt. Tutti i fotoni con energia superiore a questa vengono assorbiti e vengono trasmesse solo le radiazioni con lunghezza d'onda maggiore che gli conferiscono il caratteristico colore rosso. Il pigmento giallo cadmio, CdS, ha una banda proibita di 2,6 elettronvolt e assorbe solo la luce blu e viola. Quello che rimane della luce bianca dopo tale sottrazione presenta colore giallo. La sequenza dei colori della banda proibita, in ordine crescente della sua ampiezza, è nero, rosso, arancio, giallo, incolore.

Per quanto i semiconduttori con banda proibita molto ampia siano incolori quando sono puri, possono facilmente assumere una colorazione quando vengono «drogati» con una impurezza in tracce. L'impurezza può essere sia un donatore sia un accettore di elettroni. In entrambi i casi essa introduce un insieme di livelli energetici nella banda proibita compresa fra le bande di valenza e di conduzione. Le transizioni fra i nuovi livelli richiedono un'energia inferiore a quella necessaria per saltare l'intera banda proibita.

Nel diamante la sostituzione di pochi atomi di carbonio con altrettanti di azoto produce una banda donatrice circa 1,4 elettronvolt al di sopra della banda di valenza. L'azoto possiede un elettrone in più del carbonio ed è questo che produce la banda in questione. L'energia nominale per far saltare uno di questi elettroni nella banda di conduzione è di 4 elettronvolt, che cade ancora nell'ultravioletto. La banda è però sufficientemente ampia per assorbire un po' di luce viola. A concentrazioni di un atomo di azoto per 100 000 di carbonio questo diamante è giallo, a concentrazioni superiori diviene verde.

Il boro possiede un elettrone in meno del carbonio e origina quindi dei livelli accettori, o buche, nella banda proibita del diamante. È possibile eccitare un elettrone della banda di valenza facendogli occupare una di queste buche. Il livello dell'impurezza è centrato a 0,4 elettronvolt sopra la banda di valenza, ma si estende abbastanza in alto da consentire l'assorbimento di un po' delle radiazioni più lunghe del visibile. Il diamante drogato con boro ha quindi un colore blu.

È ben noto che il drogaggio ha influenza anche sulle proprietà elettriche dei semiconduttori, rendendo possibile la realizzazione dei moderni componenti elettronici a stato solido. Fra questi vi sono i diodi emettitori di luce e i laser a semiconduttore nei quali una corrente elettrica popola degli stati eccitati con elettroni che emettono luce quando tornano al livello fondamentale. Semiconduttori drogati possono anche comportarsi come un fosforo, sostanza che emette luce con grande efficienza quando viene eccitata elettricamente o con altri mezzi. Fosfori sono le sorgenti di luce nelle lampade fluorescenti e nei cinescopi. Sullo schermo di un cinescopio a colori sono distribuiti tre fosfori aventi emissioni nel rosso, nel blu e nel verde. Anche certi composti dei metalli di transizione si comportano come fosfori.

L'ottica fisica fornisce la spiegazione più conveniente di molti colori prodotti con meccanismi diversi che richiedono tutti una deviazione dei raggi di luce. La dispersione della luce bianca da parte di un prisma nei colori componenti avviene perché le lunghezze d'onda più piccole vengono rifratte più di quelle grandi. Similmente la diffusione della luce da parte di piccole particelle è maggiore alle piccole lunghezze d'onda e così il blu viene diffuso più del rosso. L'interferenza si osserva quando un raggio viene suddiviso in due componenti che poi vengono a riunirsi. Se le onde sono in fase quando si ricombinano l'intensità aumenta, se sono fuori fase si attenuano a vicenda. Dato che la differenza di fase fra i due fasci può dipendere dalla lunghezza d'onda, l'interferenza può, quindi, intensificare certi colori e sopprimerne altri. Nei reticoli di diffrazione la luce viene diffusa da molti centri spaziati uniformemente e i fronti d'onda multipli che ne risultano interferiscono fra di loro. Ogni lunghezza d'onda viene soppressa in certe direzioni e intensificata in altre, ottenendosi in questo modo la dispersione della luce bianca.

Ottica geometrica

Nei fenomeni descritti fino a questo punto il colore è prodotto o dall'emissione diretta di luce colorata oppure dall'assorbimento selettivo di luce. In un ultimo gruppo di fenomeni che danno origine ai colori, l'interazione della luce con la materia produce un mutamento nella direzione di propagazione della prima. Questo cambiamento di direzione è la causa prima della formazione di colori nei fenomeni di rifrazione e di diffrazione, nei quali l'ampiezza della deviazione può dipendere dalla lunghezza d'onda. Nella diffusione della luce da parte di piccole particelle la deflessione di ogni singolo raggio non è una funzione deterministica della lunghezza d'onda, ma l'intensità media della luce diffusa dipende dalla lunghezza d'onda. L'interferenza produce i colori mediante l'interazione fra raggi di luce, ma perché ciò possa avvenire è indispensabile sempre un cambiamento di direzione dei raggi.

Al livello più fondamentale questi processi possono venire compresi in termini di eccitazioni elettroniche nella materia. La rifrazione, per esempio, è dovuta a una variazione della velocità di propagazione quando la luce passa da un mezzo a un altro e la velocità di propagazione è determinata appunto dall'interazione del campo elettromagnetico della radiazione con la carica elettrica degli elettroni. Una simile analisi è sempre possibile, ma spesso è troppo laboriosa per essere veramente informativa. Occorre un'analisi a «livello superiore», che è fornita dall'ottica geometrica e dall'ottica fisica.

Un raggio di luce viene rifratto, cioè deviato, da un prisma trasparente perché la luce viene rallentata nel passaggio dall'aria al mezzo solido. (La massima velocità di propagazione della luce, una velocità che, secondo la teoria della relatività, non può mai venire superata, può venire raggiunta solo nel vuoto. In qualsiasi mezzo materiale la radiazione luminosa si propaga più lentamente.) Come una variazione di velocità possa tramutarsi in un cambiamento di direzione lo si può capire considerando un raggio di luce come una successione di fronti d'onda piani. Quando un fronte d'onda incide obliquamente su una superficie trasparente, un'estremità di esso attraversa l'interfaccia fra i due mezzi e viene rallentato prima che l'estremità opposta abbia raggiunto il nuovo mezzo di propagazione. L'estremità rallentata per prima rimane indietro rispetto al fronte d'onda rimanente e il raggio viene deviato verso la perpendicolare all'interfaccia. Quando il raggio esce da un solido trasparente si ha, invece, un allontamento dalla perpendicolare.

L'ampiezza della deviazione in seguito a rifrazione dipende dall'angolo secondo il quale il raggio di luce incide sull'interfaccia aria-solido ma, per ogni dato angolo di incidenza, dipende dal rapporto fra la velocità della luce nei due materiali. Questo rapporto, che viene detto indice di rifrazione, è generalmente diverso per le diverse lunghezze d'onda e questo produce la formazione di colori. Il rallentamento della luce in un mezzo trasparente è maggiore per le frequenze elevate (piccole lunghezze d'onda) che non per le frequenze più basse (grandi lunghezze d'onda). Ne segue che la luce viola viene rifratta più di quella rossa e quindi un raggio di luce bianca che attraversa un prisma viene disperso nei colori che lo compongono.

La dispersione della luce bianca nelle sue componenti cromatiche venne scoperta da Isaac Newton nel 1704, o almeno fu Newton che ne comprese per primo il significato. La dispersione della luce del Sole per rifrazione in goccioline d'acqua o piccoli cristalli di ghiaccio produce i colori dell'arcobaleno e di altri aloni colorati che si osservano occasionalmente intorno al Sole o alla Luna. La dispersione è anche la causa dei lampi di colore caratteristici del diamante e, in minor grado, di altre gemme trasparenti sfaccettate. La dispersione non è sempre un fenomeno gradito. Nei telescopi, negli obiettivi fotografici e negli altri sistemi ottici produce infatti l'aberrazione cromatica, ovvero la proiezione fuori registro di immagini di diversi colori.

Anche l'interferenza è stata studiata in dettaglio da Newton. La si osserva in un sistema di onde quando due di queste si combinano formando una nuova onda la cui ampiezza in ogni punto è semplicemente la somma algebrica delle ampiezze delle onde di partenza. Un'onda monocromatica può venire suddivisa in due componenti cui vengono fatti percorrere due cammini diversi prima di tornare a fonderle. Dove le due componenti sono in fase, con le creste e i ventri in coincidenza, le due onde si rafforzano e l'intensità della luce è maggiore. Dove invece le componenti sono completamente fuori fase (cresta contro ventre) si annullano e non si vede luce.

L'interferenza si può osservare spesso nei film sottili trasparenti, in cui parte della luce viene riflessa dalla prima superficie del film e parte dalla seconda. L'interferenza costruttiva o distruttiva fra i due raggi riflessi dipende dalla natura e dallo spessore del film, dall'angolo di rifrazione del raggio originale e dalla lunghezza d'onda delle radiazioni che lo compongono. Se il film ha spessore uniforme le diverse lunghezze d'onda emergono con angoli diversi. Se il film ha spessore variabile allora per ogni angolo di osservazione si vedono colori diversi in diverse posizioni. L'interferenza in una cuticola sottile e trasparente dà origine ai colori di molti insetti e farfalle. Uno strato sottile di olio sull'acqua risulta iridescente per il medesimo motivo.

Reticolo di diffrazione naturale in un opale sintetico costituito da sferette strettamente impaccate. Le sfere, costituite da biossido di silicio con un po' d'acqua, sono immerse in una matrice di composizione simile ma con indice di rifrazione leggermente diverso. Le sfere hanno un diametro di circa 250 nanometri. La dispersione della luce bianca da parte di tale reticolo tridimensionale produce i colori dello spettro che si vedono brillare nell'opale. L'opale sintetico e la fotografia al microscopio elettronico a scansione sono stati eseguiti entrambi in Francia da Pierre Gilson.

TRANSIZIONI ELETTRONICHE IN ATOMI E IONI LIBERI; TRANSIZIONI VIBRAZIONALI NELLE MOLECOLE	ECCITAZIONI ELETTRONICHE	INCANDESCENZA, FIAMME, ARCHI, SCINTILLE, FULMINI, SCARICHE NEI GAS, ALCUNI LASER
	VIBRAZIONI	COLORE VERDE-AZZURRO DELL'ACQUA PURA E DEL GHIACCIO
COLORI DEL CAMPO CRISTALLINO	COMPOSTI DI METALLI DI TRANSIZIONE	TURCHESI, MOLTISSIMI PIGMENTI, ALCUNI LASER, ALCUNI FOSFORI, ALCUNI MATERIALI FLUORESCENTI
	IMPUREZZE DI METALLI DI TRANSIZIONE	RUBINO, SMERALDO, QUARZITE ROSSA, ALCUNI LASER, ALCUNE FLUORESCENZE
	CENTRI DI COLORE	AMETISTA, QUARZO AFFUMICATO, VETRO AMETISTA DEL DESERTO, ALCUNE FLUORESCENZE
TRANSIZIONI FRA ORBITALI MOLECOLARI	TRASFERIMENTO DI CARICA	ZAFFIRO BLU, MAGNETITE
	LEGAMI CONIUGATI	COLORANTI ORGANICI, MOLTISSIMI COLORI VEGETALI E ANIMALI, LAPISLAZZULI, LUCCIOLE, LASER A COLORANTE, ALCUNE FLUORESCENZE.
TRANSIZIONI IN MATERIALI CON STRUTTURA A BANDE	CONDUTTORI METALLICI	RAME, ARGENTO, ORO, FERRO, OTTONE
	SEMICONDUTTORI PURI	SILICIO, GALENA, CINABRO, DIAMANTE
	SEMICONDUTTORI DROGATI	DIAMANTE BLU, DIAMANTE GIALLO, DIODI LUMINESCENTI, LASER A SEMICONDUTTORE, ALCUNI FOSFORI
OTTICA GEOMETRICA E FISICA	RIFRAZIONE DISPERSIVA	ARCOBALENO, «FIAMMA» NELLE GEMME, ABERRAZIONE CROMATICA
	DIFFUSIONE	BLU DEL CIELO, ROSSO DEL TRAMONTO, PIETRA DI LUNA, RUBINO ASTERIA
	INTERFERENZA	FILM DI BENZINA SULL'ACQUA, TRATTAMENTI ANTIRIFLESSI, ALCUNI COLORI DI INSETTI
	RETICOLO DI DIFFRAZIONE	OPALE, CRISTALLI LIQUIDI, ALCUNI COLORI DI INSETTI

I meccanismi di formazione del colore possono venire suddivisi in 14 categorie appartenenti a cinque classi più ampie. A parte uno solo (vibrazioni degli atomi in una molecola) tutti i meccanismi possono risalire a cambiamenti nello stato degli elettroni nella materia. Le transizioni elettroniche sono alla base di tanti meccanismi perché l'energia per eccitare un elettrone cade spesso nel campo dello spettro visibile.

Il reticolo di diffrazione

La diffrazione è dovuta alla deviazione di un'onda ai margini di un ostacolo opaco. Quando l'ostacolo è un oggetto macroscopico, l'effetto è in genere trascurabile, ma diventa importante quando le dimensioni dell'ostacolo sono poco diverse dalla lunghezza d'onda della radiazione. È quindi naturale che l'effetto dipenda fortemente dalla lunghezza d'onda.

La diffrazione è un fenomeno importante nella produzione di colori nella diffusione della luce da parte di piccole particelle. Lord Rayleigh ha mostrato che l'intensità della luce diffusa è inversamente proporzionale alla quarta potenza della lunghezza d'onda. Ne segue che la luce blu viene diffusa quattro volte di più di quella rossa. Questa differenza nella diffusione dei vari colori è evidente nel colore del cielo durante il giorno. Molecole di gas, polvere e fluttuazioni di densità dell'atmosfera diffondono particolarmente la luce blu e per questo il cielo appare di tale colore. La luce diretta del Sole, d'altro canto, viene privata di blu e appare rossastra, particolarmente all'alba e al tramonto, quando deve attraversare un maggiore spessore d'atmosferá.

Lo stesso processo si può dimostrare, su scala più piccola, facendo attraversare da un fascio di luce bianca del latte non omogeneizzato diluito con acqua, nel quale i centri di diffusione sono le particelle di grasso. Il liquido appare bluastro, mentre il fascio trasmesso direttamente diviene rossastro. La lucentezza azzurrina della pietra di luna è dovuta quasi certamente al fenomeno della diffusione. La diffusione da parte di inclusioni di dimensioni relativamente grandi, che non alterano il colore, spiega il riflesso caratteristico (gatteggio) del rubino e dello zaffiro asteria e del quarzo occhio di tigre.

Interferenza e diffusione diffrattiva si combinano nell'ultimo sistema di generazione del colore: il reticolo di diffrazione. Questo reticolo è costituito da una successione di molte linee o punti equidistanti, con spaziatura non troppo grande in rapporto alla lunghezza d'onda della luce. La luce che attraversa il reticolo viene diffusa in tutte le direzioni in corrispondenza di ogni fenditura e così le onde che passano attraverso fenditure adiacenti possono interferire. Per ogni lunghezza d'onda esiste un angolo rispetto al fascio incidente per cui l'interferenza è costruttiva. Per ogni altro angolo le onde si sottraggono. L'ampiezza di tali angoli dipende solo dalla lunghezza d'onda e dalla distanza fra le righe o punti del reticolo. Quando un fascio di luce bianca attraversa un reticolo o viene riflessa da questo, ogni lunghezza d'onda subisce interferenza costruttiva per un diverso insieme di direzioni. Quindi la luce in ogni direzione è monocromatica ma, ruotando il reticolo si osserva una serie di spettri, in parte intersecantesi.

Un reticolo di diffrazione può venire costruito incidendo una serie di righe parallele su una lastra di vetro, ma un certo numero di sistemi naturali possiedono pure le dimensioni e l'elevato grado di ordine necessari per ottenere un reticolo efficiente nel visibile. Per esempio, nelle

sostanze che danno origine ai cosiddetti cristalli liquidi le molecole si trovano impilate con sufficiente regolarità per comportarsi come un reticolo di diffrazione. Dato che la distanza media fra le molecole dipende dalla temperatura, anche il colore del materiale è funzione di questa e può infatti venire impiegato come termometro ottico. Simili reticoli ordinati di molecole danno origine a colori di diffrazione in certi insetti e certe farfalle. Colori spettrali dovuti a diffrazione si osservano anche quando si vede un lampione stradale attraverso il tessuto di un ombrello o si osserva un disco fonografico sotto illuminazione radente.

Il più caratteristico reticolo di diffrazione naturale è costituito dall'opale. Questa pietra è costituita da sfere di biossido di silicio e acqua strettamente e ordinatamente impaccate in un reticolo tridimensionale con una spaziatura di 250 nanometri. La matrice trasparente o translucida che riempie gli spazi tra le sfere ha una composizione chimica simile, ma un indice di rifrazione leggermente diverso. Quando un raggio di luce bianca incide su questo reticolo di diffrazione tridimensionale, all'interno della pietra compaiono dei colori puri che mutano quando si muove l'occhio o la pietra stessa. Il meccanismo secondo il quale vengono prodotti i colori dell'opale è stato compreso solo nel 1964, quando la struttura di questo minerale poté venire risolta al microscopio elettronico. Poco tempo dopo fu possibile produrre degli opali sintetici.

Lo spettro visibile

In questo catalogo, che certamente è incompleto, ho descritto più di una dozzina di meccanismi di formazione del colore. Essi possono venire riuniti in cinque classi: eccitazioni di atomi o ioni liberi e vibrazioni di molecole, effetti del campo cristallino, transizioni di orbitali molecolari, transizioni nelle bande di solidi ed effetti interpretabili mediante l'ottica fisica. Può sembrare una coincidenza straordinaria che tanti fenomeni così diversi avvengano in una banda di lunghezze d'onda con un'ampiezza di meno di un'ottava. E può sembrare ancora più incredibile che tale banda ristretta coincida con quella cui è sensibile l'occhio umano.

In realtà questa può benissimo non essere una coincidenza. L'interesse di questa piccola regione dello spettro è dovuto semplicemente al fatto che queste sono le lunghezze d'onda in corrispondenza delle quali cominciano a diventare importanti le interazioni fra gli elettroni e la luce. Radiazioni di minore energia possono influenzare il movimento di atomi e di molecole e vengono quindi percepite come calore. Radiazioni di energia maggiore possono ionizzare gli atomi e danneggiare permanentemente le molecole, cosicché i loro effetti sono per lo più distruttivi. Solo in questa piccola zona di transizione fra i due estremi si osserva una buona sintonia fra le energie della radiazione visibile e quelle della struttura elettronica della materia.

Gli Appalachi meridionali e la crescita dei continenti

Il rilevamento sismico a riflessione fa pensare che per almeno metà della storia terrestre i continenti si siano evoluti per accavallamento e rimescolamento di strati sottili di materiale posto ai loro margini

di Frederick A. Cook, Larry D. Brown e Jack E. Oliver

La teoria della tettonica a zolle, secondo la quale la crosta terrestre è la parte superiore di una serie di zolle rigide interagenti, ha dato un notevole contributo alla descrizione dell'evoluzione dei fondi oceanici, ma non ha chiarito molte cose a proposito della formazione e della struttura dei continenti. La situazione sta incominciando a cambiare con l'applicazione intensiva allo studio della crosta profonda e del sottostante strato superiore del mantello della tecnica esplorativa, tipica dell'industria petrolifera, della prospezione a riflessione sismica. Questa tecnica, che si basa sullo studio delle onde elastiche riflesse da discontinuità nella densità delle rocce e nella velocità di propagazione attraverso la roccia stessa, ha permesso di fissare sulla carta nuovi particolari della struttura geologica del basamento continentale.

L'applicazione di questa tecnica alla zona degli Appalachi meridionali, negli Stati Uniti, rivela il modo in cui i continenti si modificano a mano a mano che i bacini oceanici si chiudono e i continenti entrano in collisione in corrispondenza delle zone di subduzione, costituite dalle profonde fosse oceaniche vicino ai margini continentali dove la crosta terrestre affonda nel mantello. Da molto tempo si sospetta che la crescita dei continenti sia dovuta all'aggiunta di materiale crostale in corrispondenza delle zone di subduzione. I profili ottenuti con la riflessione sismica degli Appalachi meridionali e studi geologici connessi rivelano che è proprio così che vanno le cose: i continenti si evolvono per effetto dell'accavallamento e del rimescolamento di sottili strati orizzontali di materiale crostale.

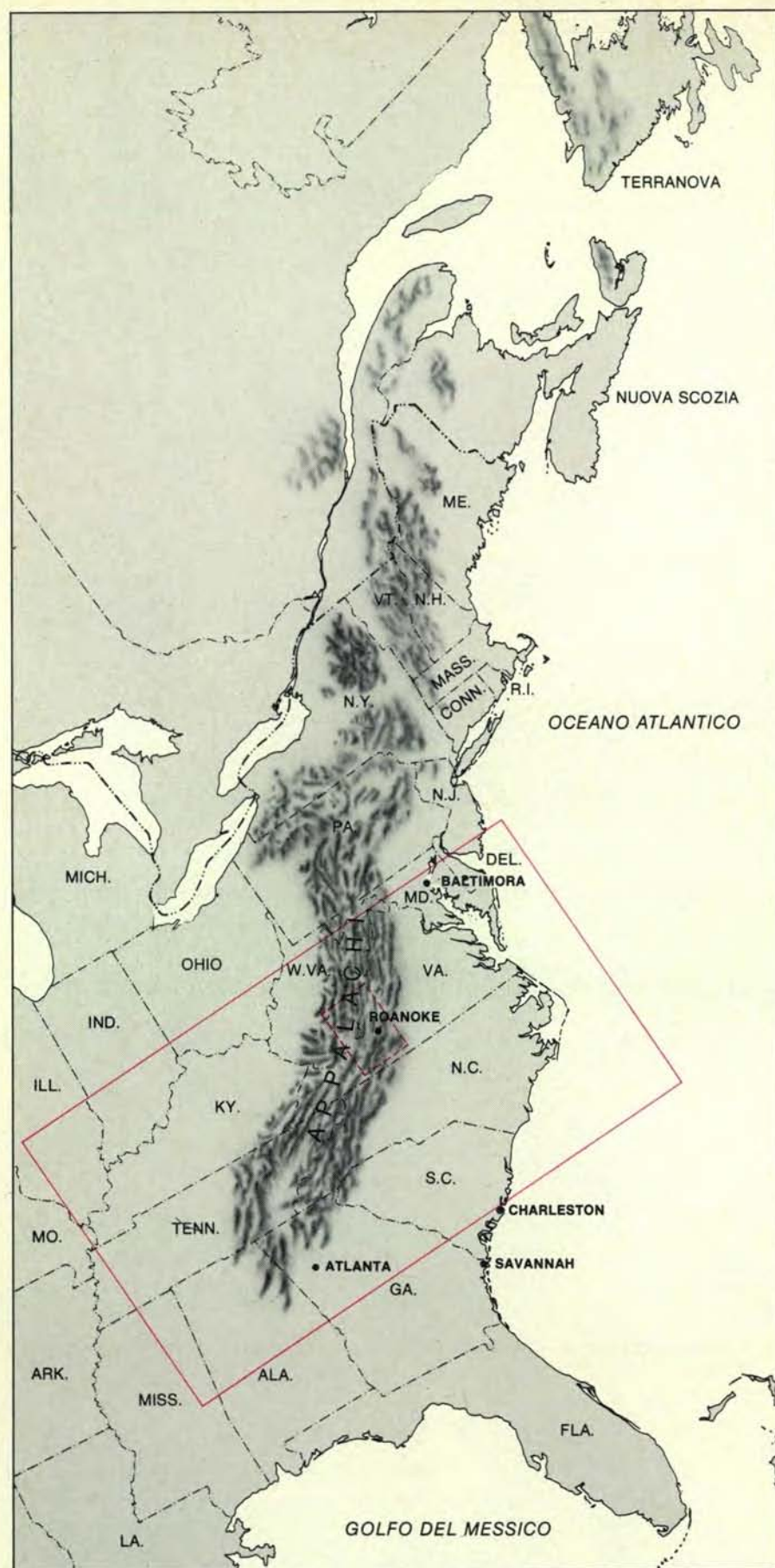
Le prospezioni sismiche sono state compiute dal Consortium for Continental Reflection Profiling (COCORP), un gruppo di geologi e geofisici provenienti dalle università, dall'industria e dal governo con a capo Sidney Kaufman, della Cornell University, e due di noi (Brown e Oliver). Il COCORP ha raccolto i suoi primi dati a Hardeman County (Texas) nel 1975. Dopo di che le prospezioni si sono svolte in 11 siti. Sondando le rocce cristalline sotto strati sedimentari fino a profondità dell'ordine di 50 chilometri, lo studio si è concentrato sul basamento continentale, e ha avuto successo. Il COCORP ha tracciato il profilo di una massa mediocrostale di magma nel Nuovo Messico centrale, ha esplorato un'antica *rift valley* sepolta nel Michigan e ha seguito le tracce, fino nelle profondità della crosta, di una importantissima faglia di compressione nel Wyoming sudoccidentale. Prima ancora che l'indagine esplorativa del Wyoming fosse condotta a termine, erano già sorte molte controversie a proposito della profondità e della giacitura della faglia, la quale delimita un'importante caratteristica morfologica delle Montagne Rocciose.

La scoperta più spettacolare è stata fatta negli Appalachi meridionali, dove i profili hanno rivelato che quelle montagne sono sostenute, fino a una profondità di almeno 18 chilometri, da strati orizzontali di materiale che è, o lo era una volta, sedimentario. Le rocce sedimentarie sono come è noto uno dei tre tipi di roccia, mentre gli altri due sono costituiti dalle rocce ignee e da quelle metamorfiche. Le rocce sedimentarie hanno origine da sedimenti depositati sulla superficie dall'acqua e dal vento. Le rocce ignee si formano dal magma che si raffredda e si indurisce. Le rocce metamorfiche si formano quando quelle sedimentarie o ignee sono sottoposte a calore o a pressione per lunghi periodi di tempo. Le rocce metamorfiche si possono definire di alto o basso metamorfismo a seconda della misura in cui sono state ricristallizzate nel processo di metamorfismo.

Buona parte di quelle che si trovano sulla superficie degli Appalachi meridionali sono rocce metamorfiche estremamente deformate. Inoltre sono più vecchie o contemporanee degli strati sedimentari orizzontali che sono stati scoperti sotto di esse. Questo fatto fa pensare che circa 475 milioni di anni or sono le rocce superficiali incominciarono a essere trasportate in forma di faglia sottile per almeno 260 chilometri sopra il margine continentale orientale della massa di terra che doveva diventare l'America Settentrionale. Questa scoperta conferma l'ipotesi secondo la quale lo scorrimento orizzontale o quasi può trasportare a grandi distanze grandi quantità di materialè crostale e rende plausibile supporre che un continente abbia potuto accrescersi ed evolversi per la dislocazione di sottili porzioni orizzontali di materiale in corrispondenza dei margini continentali.

Non è da escludere che i profili di riflessione sismica abbiano anche un'importanza di carattere più immediato, in quanto hanno indotto i geologi a ipotizzare l'esistenza di depositi sconosciuti di petrolio o di gas nelle rocce sedimentarie sotto una parte delle faglie di sovrascorrimento. Le rocce sedimentarie sono in particolare il migliore obiettivo delle ricerche petrolifere e gassifere perché, a differenza delle rocce ignee e della maggior parte di quelle metamorfiche, esse non sono state sottoposte a pressioni e a temperature tali da distruggere o espellere gli idrocarburi.

I primi studi compiuti con la riflessione sismica dei bacini sedimentari risalgono agli anni venti. Da allora la tecnica è stata sviluppata in maniera intensiva

Immagine a falsi colori degli Appalachi a ovest di Roanoke (Virginia) ripresa in autunno da un satellite. Roanoke è l'area in color grigio chiaro a destra sotto la parte centrale dell'immagine. Il rosso indica le zone coperte da vegetazione, mentre le superfici d'un vivo color verde bluastro sono fiumi e laghi. Qui le montagne, che raggiungono altitudini comprese fra 500 e 1000 metri, seguono un andamento che prima è in direzione nord-est e poi si sposta più verso nord. Le direttrici lineari costituiscono il materiale sedimentario della provincia di Valley and Ridge.

Questa cartina degli Stati Uniti orientali e del Canada meridionale mostra come gli Appalachi si estendano dall'Alabama centrale a Terranova. Il rettangolo tratteggiato delimita la zona di Roanoke presentata nell'immagine della pagina precedente ripresa da un satellite. Il rettangolo in colore delimita invece la zona degli Appalachi meridionali riportata a pagina 60.

dall'industria petrolifera nel quadro dei suoi costanti tentativi di localizzare formazioni petrolifere e gassifere di importanza economica. La teoria di questa tecnica è semplicissima. Le onde elastiche prodotte da un'esplosione o da qualche altra sorgente di energia alla superficie si propagano nel sottosuolo finché non incontrano delle discontinuità negli strati di roccia. Le onde vengono riflesse dovunque vi sia un mutamento improvviso della densità della roccia o della velocità delle onde. L'energia riflessa viene rivelata alla superficie da una serie di geofoni, o sensori di onde sonore.

Il tempo che le onde impiegano per arrivare dalla sorgente alla discontinuità riflettente e tornare di nuovo alla superficie dipende dalla profondità della discontinuità. Ulteriori indicazioni sulle discontinuità possono essere tratte dalle registrazioni di una serie di geofoni che rispondono a onde il cui percorso è quasi verticale. I profili di riflessione sismica usati per localizzare depositi di petrolio e di gas prendono in considerazione solo le onde con tempi di percorrenza inferiori a quattro o cinque secondi, che corrispondono a profondità comprese fra otto e dieci chilometri. Le onde con tempi di percorrenza superiori vengono riflesse da rocce troppo profonde per poter essere raggiunte con le attuali tecniche di estrazione del petrolio. Il COCORP ha lavorato con onde con tempi di percorrenza dell'ordine di 20 secondi, corrispondenti a profondità di 60-70 chilometri.

Il COCORP ha compiuto un rilevamento a riflessione pluricanale con ben 2304 geofoni (suddivisi in 96 gruppi di 24 unità) disposti su un tratto di oltre sei chilometri e mezzo. Tutta la schiera dei rilevatori era collegata per mezzo di un cavo multipolare a un autocarro in cui i dati raccolti venivano registrati su nastro magnetico sotto forma numerica per essere elaborati successivamente da calcolatori che avrebbero tracciato un profilo sismico della geologia sotterranea. Se le sorgenti delle onde e i geofoni vengono disposti nella maniera più opportuna e i dati vengono elaborati correttamente, ogni disturbo indesiderabile verrà eliminato e i segnali riflessi risulteranno intensificati.

Sebbene nelle prospezioni petrolifere e gassifere a volte si faccia ancora uso di esplosivi come sorgente arficiale di energia elastica, il COCORP, come buona parte dell'industria petrolifera, si è servito di una sorgente non esplosiva, accettabile sul piano ambientale, chiamata Vibroseis, che fu inventata negli anni cinquanta dalla Continental Oil Company. In un tipico rilevamento COCORP, quattro o cinque vibratori installati su un autocarro dirigono nel terreno un segnale compreso fra 8 e 32 hertz. Nel periodo di emissione di 30 secondi la frequenza viene fatta variare linearmente in modo da formare un «cinguettio». Dopo ogni emissione i vibratori vengono spostati avanti di qualche metro e viene fatta un'altra emissione. Dopo 16 emissioni che coprono un tratto di 122 metri, il processo viene ripetuto alla successiva stazione della sorgente, distante 134 metri.

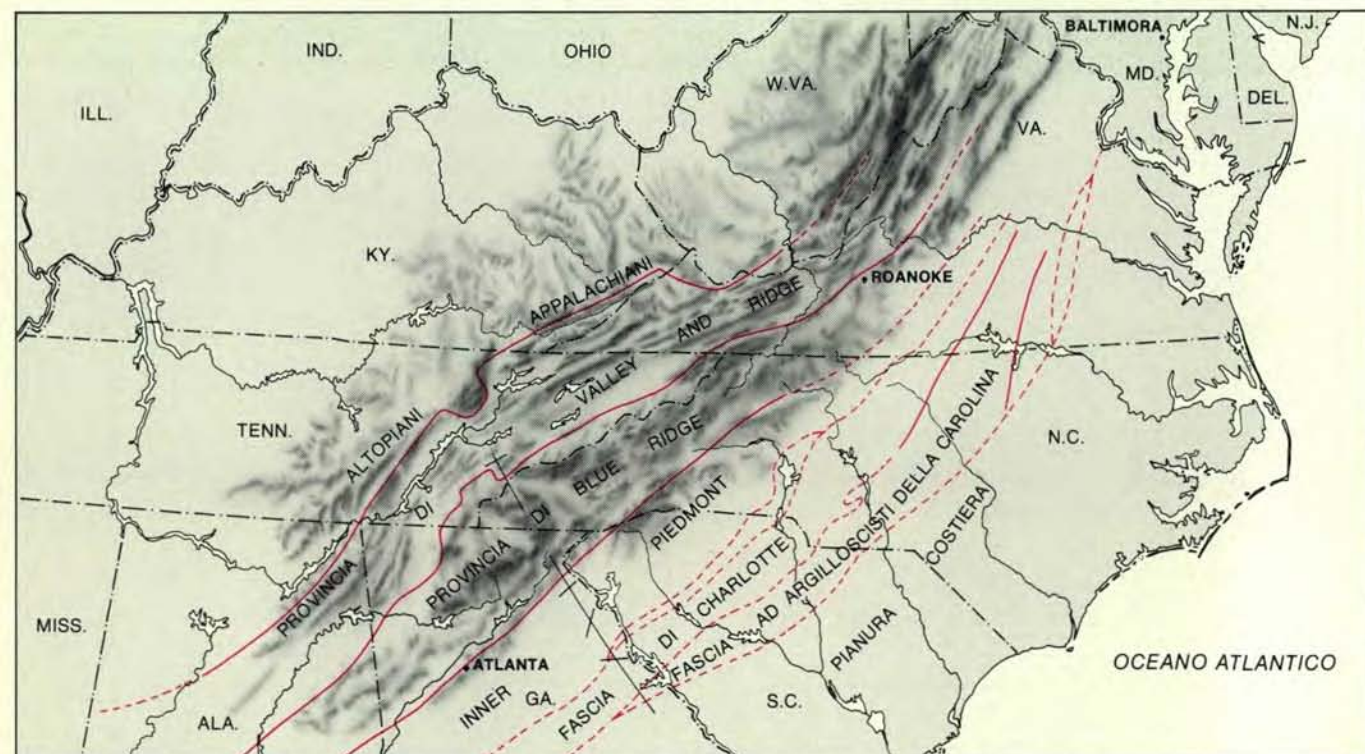
Tutto questo ha lo scopo di registrare molte riflessioni provenienti da uno stesso punto del sottosuolo. La sovrabbondanza dei dati permette all'analisi fatta al calcolatore di sopprimere i disturbi di origine esterna, di ampliare la forza del segnale riflesso e di valutare la velocità delle onde vibrazionali sotterranee. Sulla scorta dei dati raccolti da molte e diverse configurazioni di sorgenti e di sensori viene poi tracciato un profilo sismico della struttura del basamento continentale. Il procedimento assomiglia un po' alla tecnica medica della tomografia assiale computerizzata, in cui le radiografie ai raggi X prese da molte angolazioni diverse vengono combinate in modo da produrre una rappresentazione della struttura interna del corpo in sezione trasversale.

Nella prospezione a riflessione sismica ci vogliono spesso alcuni mesi per raccogliere ed elaborare i dati relativi a un particolare sito. Le sorgenti e i rivelatori vengono spostati sulla superficie alla media di uno-quattro chilometri al giorno. Per ricostruire la struttura geologica del sottosuolo è necessaria l'elaborazione di una quantità enorme di dati. Per esempio, l'analisi di un'importante trasversale nord-ovest-sud-est nella Georgia e nel Tennessee ha richiesto la manipolazione di circa due miliardi di dati, valore ottenuto da 3843 punti di vibrazione moltiplicato per 96 canali moltiplicato per un tempo di registrazione di 50 secondi diviso per un quoziente campione di 0,008 secondi.

L'analisi al calcolatore dei dati raccolti dal COCORP comprendeva la «demultiplazione» dei segnali registrati attraverso il cavo multicanale, l'eliminazione di dati particolarmente rumorosi, il calcolo delle velocità delle onde, la raccolta di tutti i segnali riflessi da ogni punto comune di profondità, la compensazione della differenza di distanza fra le sorgenti e i ricevitori, la regolazione delle differenze in fatto di topografia e di geologia vicino alla superficie e la sovrapposizione dei molti segnali coerenti per ogni punto comune di riflessione. Il risultato di questa elaborazione intensiva è uno spaccato sismico analogo a uno spaccato geologico, salvo due importanti differenze: primo, che la profondità è rappresentata non dalla distanza, ma dal tempo di percorrenza e, secondo, che le variazioni laterali della velocità delle onde nelle rocce possono distorcere la geometria delle riflessioni.

Per desumere un preciso spaccato geologico dallo spaccato sismico sono necessarie notevole abilità ed esperienza. Le interfacce inclinate, per esempio, sono rappresentate sullo spaccato sismico da riflessioni che sembrano fuori posizione; quanto più ripida è la pendenza, tanto più grande è la «collocazione errata». Spesso per avere un quadro preciso dell'effettiva geologia del sottosuolo è necessario far «migrare» a ritroso le onde riflesse.

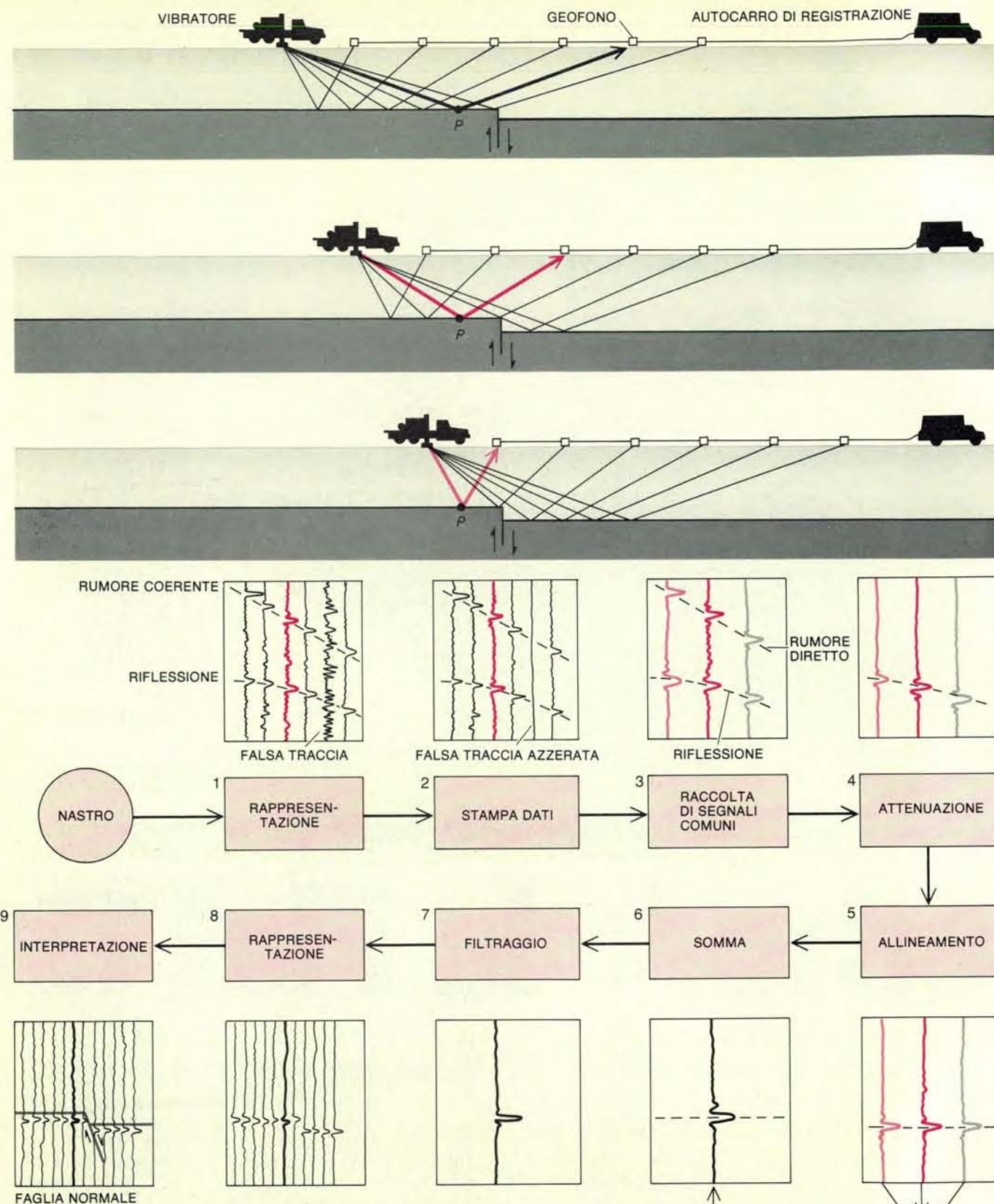
Questa è stata la tecnica esplorativa applicata agli Appalachi meridionali. Gli Appalachi si estendono da Terranova all'Alabama centrale per più di 3000 chilometri. Gli Appalachi meridionali sono costituiti da una serie di province geologiche ben distinte e di *belt* (fascia) orientate da nord-est a sud-ovest. («Provincia» è un termine generico usato per indicare qualsiasi zona le cui rocce hanno una storia analoga, mentre con il termine «fascia» si vuole invece indicare una struttura geologica lunga e lineare le cui rocce hanno una composizione analoga.) Da nord-ovest a sud-est gli Appalachi meridionali sono composti dalla provincia di Valley and Ridge, dalla provincia di Blue Ridge, dalla provincia di Piedmont (che comprende l'Inner Piedmont, la fascia di Charlotte e la fascia di argilloscisti della Carolina) e dalla pianura costiera.

La provincia di Valley and Ridge è caratterizzata da strati di pieghe e faglie di compressione di rocce sedimentarie in prevalenza non metamorfosate, formatesi fra 600 e 300 milioni di anni fa. Le faglie di compressione e le pieghe indicano che le rocce erano molto compresse in senso orizzontale. Per molto tempo non si ebbe modo di sapere se la deformazione della provincia di Valley and Ridge coinvolgesse anche le rocce cristalline che formano il basamento sotto gli strati sedimentari (nel qual caso il processo sarebbe stato chiamato «tettonica da movimento del basamento») o se la deformazione si limitasse principalmente ai sottili strati sedimentari sopra il basamento. Oggi invece, grazie ai profili di riflessione sismica e alle prospezioni effettuate per la ricerca di giacimenti di petrolio e di gas, sappiamo che la deformazione riguarda gli strati sedimentari sottili. Simili a un tappeto corrugato sul pavimento, gli strati sedimentari sembrano essersi spostati verso occidente, al di sopra del basamento cristallino, appoggiati su grandi zone orizzontali di separazione.

Gli Appalachi meridionali sono formati da quattro caratteristiche topografiche disposte da nord-ovest a sud-est: la provincia di Valley and Ridge, la provincia di Blue Ridge, la provincia di Piedmont (che comprende l'Inner Piedmont, la fascia di Charlotte e la fascia ad argilloscisti della Carolina) e la pianura costiera. Il COCORP ha studiato il basamento continentale lungo la linea in nero. Le linee in rosso indicano faglie di compressione.

VIBRATORE GEOFONO AUTOCARRO DI REGISTRAZIONE

RUMORE COERENTE RIFLESSIONE FALSA TRACCIA FALSA TRACCIA AZZERATA RUMORE DIRETTO RIFLESSIONE

NASTRO → 1 RAPPRESENTAZIONE → 2 STAMPA DATI → 3 RACCOLTA DI SEGNALI COMUNI → 4 ATTENUAZIONE → 5 ALLINEAMENTO → 6 SOMMA → 7 FILTRAGGIO → 8 RAPPRESENTAZIONE → 9 INTERPRETAZIONE

FAGLIA NORMALE

Rappresentazione estremamente schematizzata delle nove fasi dell'elaborazione dei dati. Vibratori montati su un autocarro *(in alto)* dirigono nel terreno un segnale a bassa frequenza che verrà parzialmente riflesso dalle superfici di discontinuità della roccia. L'energia riflessa viene rilevata alla superficie da una serie di geofoni che sono collegati a un autocarro in cui i dati vengono raccolti e memorizzati. La loro successiva elaborazione rivelerà la geologia del sottosuolo. Quindi i vibratori vengono spostati avanti di qualche metro. Le linee spesse in nero e in colore indicano il percorso di un segnale riflesso da un punto *P* per tre diverse posizioni di un vibratore. L'elaborazione dei dati avviene in nove fasi *(in basso)*. Dai segnali registrati *(1)* vengono eliminati i dati di disturbo *(2)*. Poi si raccolgono i segnali delle riflessioni provenienti da un punto comune di profondità *(3)*. Attenuati i disturbi d'ampiezza elevata *(4)*, i segnali comuni vengono allineati *(5)*, in modo da essere sommati in fase *(6)*. L'impulso sommato viene filtrato *(7)* e vengono presentati i segnali sommati *(8)*. Si procede in fine all'interpretazione dei dati *(9)*.

A sud-est della provincia di Valley and Ridge c'è la provincia di Blue Ridge. La principale transizione fra l'una e l'altra è una grande faglia di compressione digradante verso sud-est. A differenza delle rocce sedimentarie della provincia di Valley and Ridge, le rocce della provincia di Blue Ridge sono state sottoposte in genere a un metamorfismo particolarmente intenso. In entrambe le province, però, le rocce sono state deformate in tre orogenesi o episodi di sollevamento di catene montuose: la Taconica, l'Acadiana e l'Allegheniana.

Il basamento della provincia di Blue Ridge comprende rocce precambriane che hanno almeno un miliardo di anni. Molti geologi hanno citato questo fatto come prova che la provincia di Blue Ridge è radicata sul posto, asserendo che essa fa parte del corpo principale roccioso del basamento che in quel punto forma la crosta. Nuovi dati indicano, peraltro, che, sebbene in quella zona le rocce sedimentarie non metamorfosate siano rare, esse si trovano occasionalmente in alcune «finestre»: aree in cui le rocce metamorfiche sono state erose in modo da scoprire il materiale sottostante. Le finestre si trovano per lo più nella Blue Ridge occidentale, e così oggi sappiamo che almeno lì le rocce cristalline del basamento sono sostenute da strati di rocce sedimentarie non metamorfosate. Inoltre, una delle finestre più grandi, quella della Grandfather Mountain della North Carolina, si trova nella Blue Ridge orientale. Quest'area, che mette allo scoperto rocce carbonatiche e scisti cambriani, dimostra che tanto sotto la Blue Ridge orientale quanto sotto quella occidentale sono presenti rocce sedimentarie che sono state sottoposte a debole metamorfismo.

Al confine orientale della provincia di Blue Ridge c'è una caratteristica topografica di grande rilievo, nota col nome di zona di Brevard, che si estende dall'Alabama alla Virginia. Si tratta di una fascia ristretta di rocce plurideformate che segna il confine tra la Blue Ridge e la Piedmont. La storia e la natura della deformazione della zona di Brevard sono state oggetto di molte discussioni, tanto che il primo rilevamento compiuto dal COCORP nella zona era dettato dalla speranza di ricavare nuovi dati in modo da risolvere la controversia. Una delle scoperte geologiche più importanti nella zona di Brevard è stata fatta da Robert D. Hatcher, Jr., dell'Università della South Carolina, il quale ha interpretato la presenza di rocce sedimentarie molto poco metamorfosate come prova dell'esistenza di strati sedimentari in profondità sotto la zona in questione e ha asserito che le rocce erano state portate alla superficie dal processo di fagliazione. Questo costituiva una prova ulteriore della possibile presenza sotto la Blue Ridge di materiale sedimentario che si prolunga verso est almeno fino alla zona di Brevard.

A sud-est di questa zona c'è l'Inner Piedmont, costituito principalmente da rocce di alto metamorfismo, in cui si sono

intrusi corpi di rocce ignee come i graniti della Stone Mountain e quelli di Elberton. Del Piedmont si è sempre pensato tradizionalmente che fosse il «nucleo» metamorfico degli Appalachi meridionali. Poiché pare che buona parte del materiale metamorfico sia costituito da rocce sedimentarie deformate molto estesamente, i geologi pensarono che il Piedmont fosse stato sollevato in senso verticale e che la deformazione si fosse propagata verso occidente nella Blue Ridge e nella Valley and Ridge. I nuovi dati sismici hanno completamente modificato quest'opinione.

L'Inner Piedmont è fiancheggiato a sud-est dalla fascia della Kings Mountain, una stretta banda di rocce sedimentarie metamorfosate e di rocce vulcaniche, alcune delle quali potrebbero essere i resti di un oceano chiuso o di un bacino geologico in corrispondenza del margine continentale. Le pieghe di maggior rilievo di questa fascia e le pieghe dell'area a sud-est sono allineate in senso nord-est, analogamente alle pieghe dell'Inner Piedmont e della Blue Ridge. In questa fascia vi sono i segni di almeno due periodi di deformazione, di cui uno risale probabilmente a circa 450 milioni e l'altro a circa 350 milioni di anni fa.

Il Piedmont sudorientale è formato dalla fascia di Charlotte, che è costituita da rocce sedimentarie metamorfosate, e dalla fascia ad argilloscisti della Carolina, che consiste soprattutto di materiale vulcanico metamorfosato. Si ha l'impressione che questa fascia sia tutto ciò che rimane di un arco vulcanico simile a quelli che oggi sono attivi nel Pacifico occidentale. Qui il vulcanismo incominciò probabilmente nel tardo Precambriano, fra 700 e 650 milioni di anni fa, e proseguì nel Cambriano fino a circa 500 milioni di anni fa.

La pianura costiera a sud-est del Piedmont consiste di una sequenza di sedimenti giovani (meno di 200 milioni di anni di età) che coprono un basamento cristallino. Pozzi scavati fino a penetrare nel basamento hanno portato alla luce rocce sedimentarie metamorfosate simili a quelle dell'Inner Piedmont e rocce vulcaniche metamorfosate simili a quelle della fascia ad argilloscisti della Carolina. Forse il basamento vulcanico e sedimentario metamorfosato del Piedmont e della fascia ad argilloscisti della Carolina si estende sotto la pianura costiera.

Secondo il quadro storico rivelato dalla teoria della tettonica a zolle, 200 milioni di anni fa i continenti che ora delimitano l'Atlantico erano uniti fra loro come i pezzi di un rompicapo a incastro in modo da formare un'unica immensa massa di terraferma. A quell'epoca l'America Settentrionale incominciò a separarsi dall'Europa, dall'Africa e dall'America Meridionale. Poiché un tempo questi continenti erano uniti, non è da escludere che i dati geologici della parte orientale dell'Atlantico settentrionale aiutino a spiegare la struttura e la formazione degli Appalachi. Basandosi sulla teoria della tettonica a zolle, quasi tutti i modelli del periodo Carbonifero prima che l'Atlantico incominciasse ad aprirsi collocano l'Africa occidentale adiacente alla parte meridionale dell'America Settentrionale, ma una recente interpretazione di dati paleomagnetici da parte di Edward Irving, del Department of Energy, Mines and Resources del Canada, fa pensare che fosse la parte settentrionale dell'America Meridionale a essere adiacente a quell'epoca alla parte orientale dell'America Settentrionale. In ogni caso l'Africa occidentale e la parte settentrionale dell'America Meridionale hanno entrambe delle fasce di corrugamento e di compressione che probabilmente si formarono nello stesso episodio di orogenesi del Carbonifero.

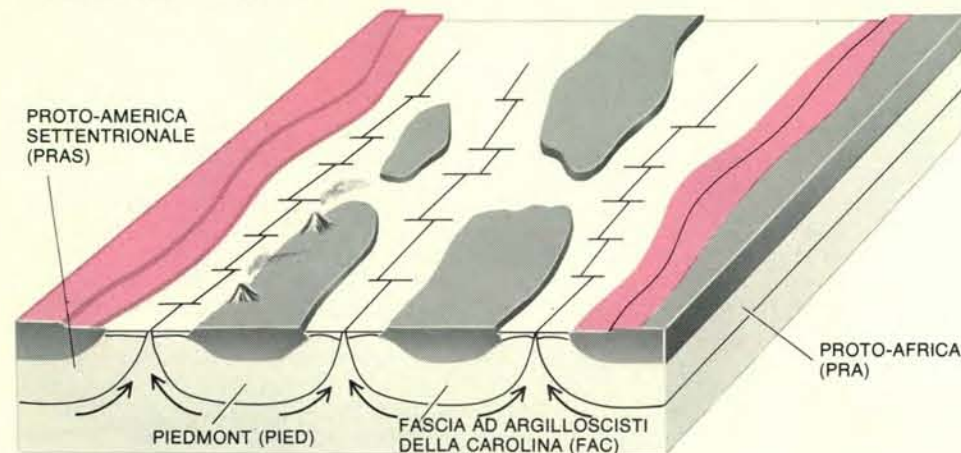
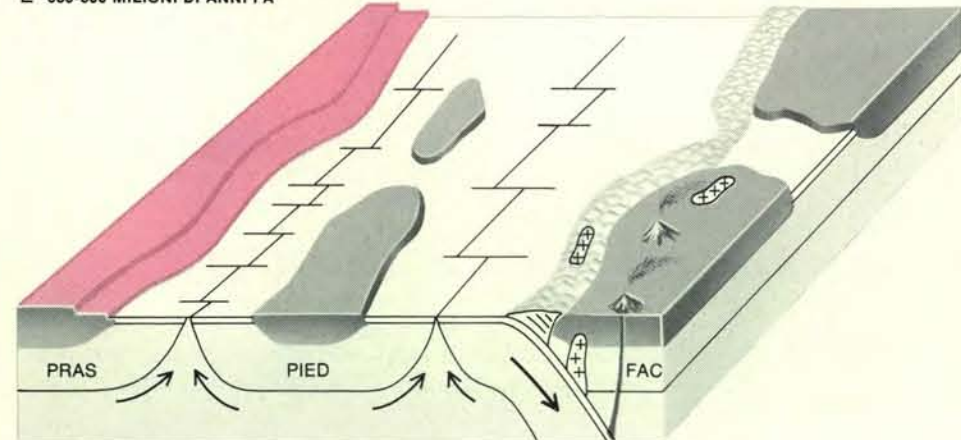
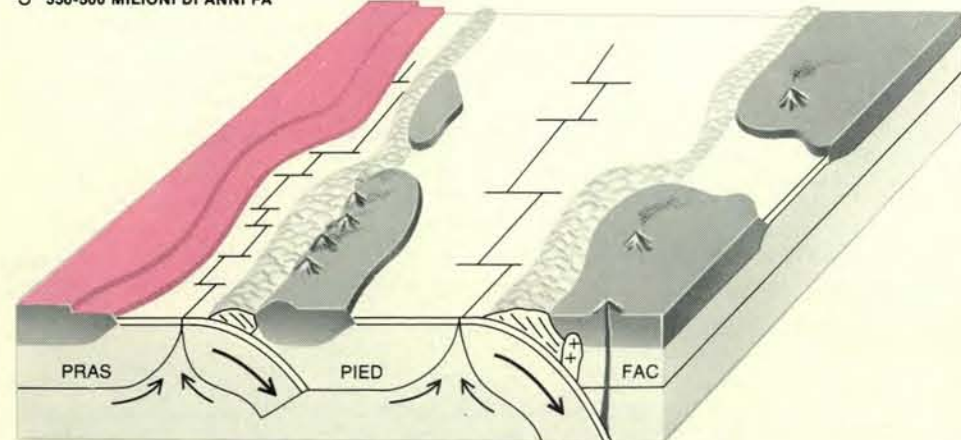
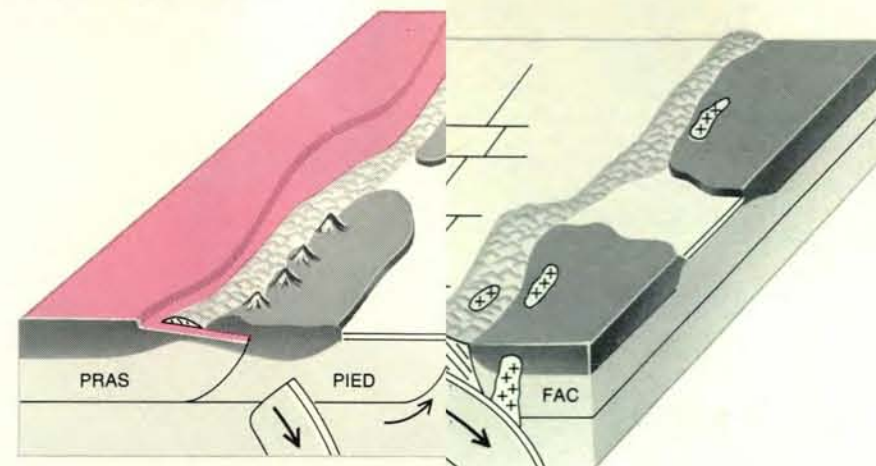
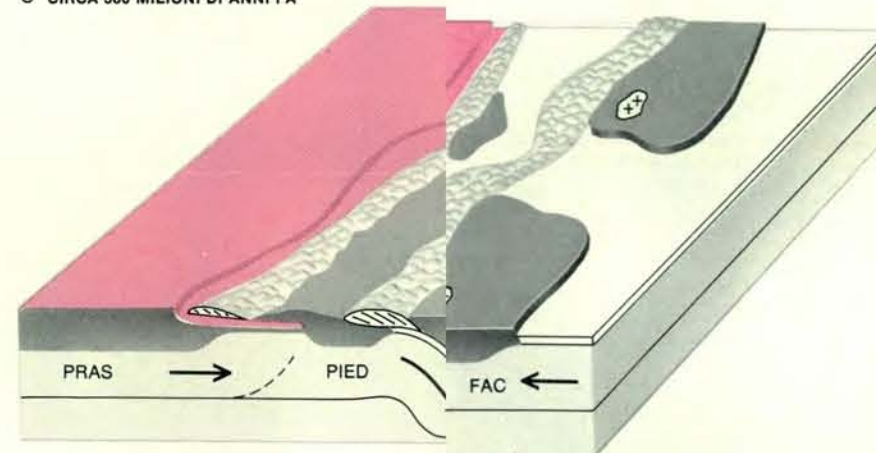
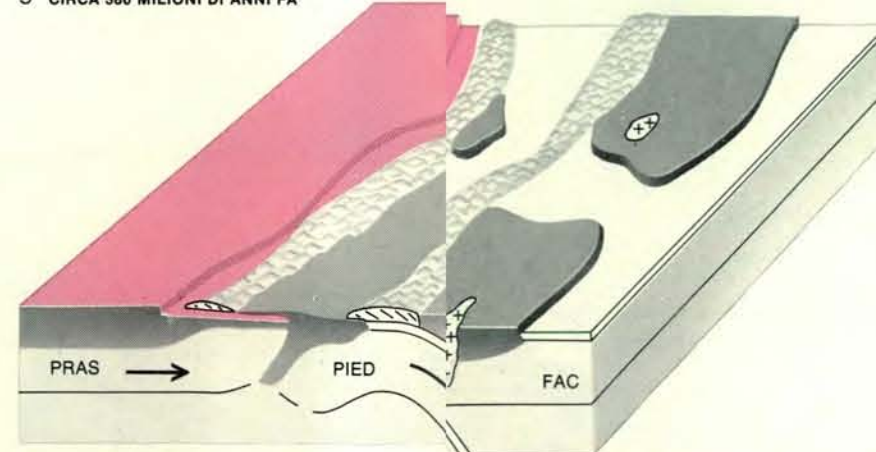
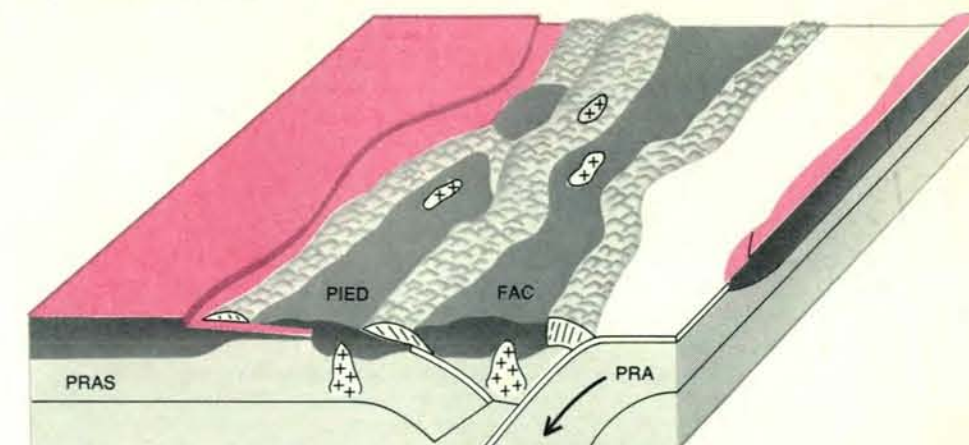
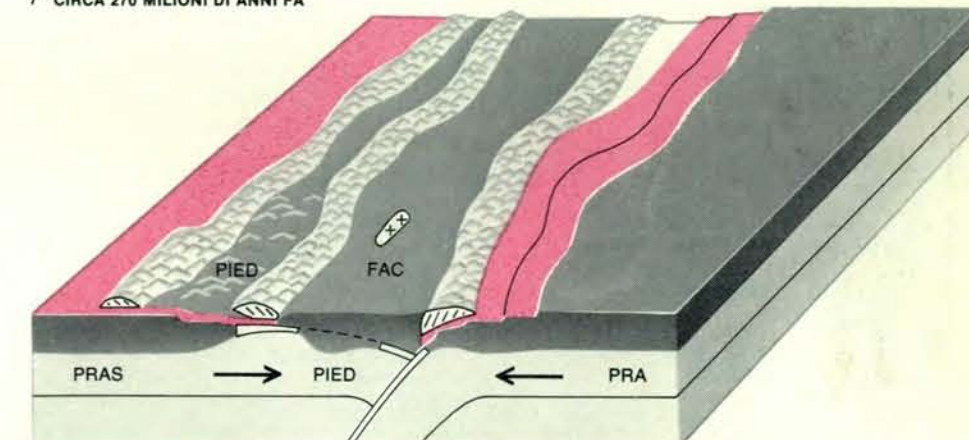
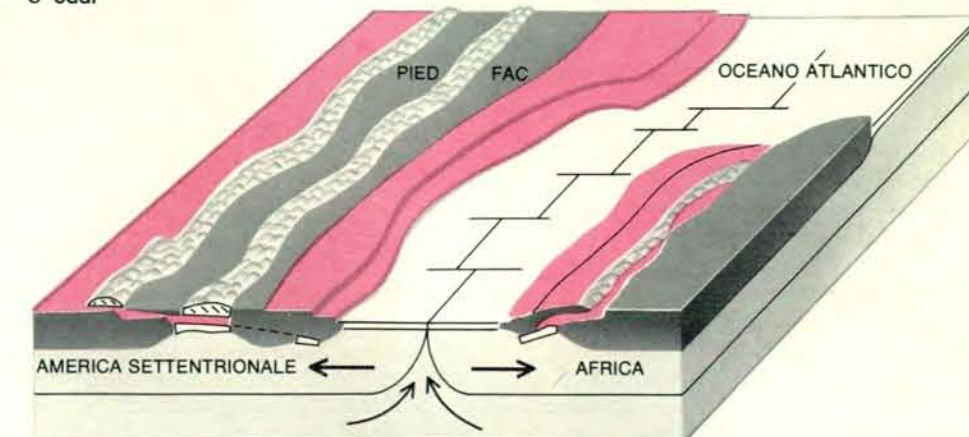
La catena montuosa delle Mauritanidi dell'Africa occidentale è caratterizzata da est a ovest da una serie di fasce che assomigliano in qualche modo a quelle appalachiane. Le Mauritanidi orientali sono composte di strati sedimentari non metamorfosati parzialmente coperti da rocce metamorfiche che si sono sovrapposte a sedimenti provenienti da ovest lungo faglie di compressione. A ovest vi sono rocce metamorfiche di qualità superiore, più vecchie, che assomigliano a quelle del Piedmont appalachiano meridionale. Una pianura costiera di rocce orizzontali più giovani copre il resto dell'orogene. Una mappa geologica della zona fa supporre che circa 550 milioni di anni fa un lieve episodio di deformazione sia stato seguito da un periodo di metamorfismo e di scorrimento prima dell'apertura dell'Atlantico. Questo periodo probabilmente corrisponde all'orogenesi Allegheniana negli Appalachi. In senso lato, le Mauritanidi dell'Africa occidentale sono un'immagine speculare degli Appalachi.

Il COCORP ha studiato a fondo gli Appalachi nel Tennessee, nella North Carolina e nella Georgia. Il rilevamento iniziale copriva una zona che andava da un centinaio di chilometri a sud-est di Knoxville (Tennessee) a un centinaio di

Rappresentazione estremamente schematizzata del profilo di riflessione sismica degli Appalachi meridionali *(in alto)*. Sull'asse verticale sono indicati i tempi di andata e ritorno delle onde riflesse. L'asse orizzontale localizza le più importanti caratteristiche geologiche e indica la distanza, a livello di superficie, dall'estremità nordoccidentale dell'inizio della zona di rilevamento del COCORP. Sono segnati anche i confini tra i vari stati. Sotto il profilo schematico c'è una fotografia del profilo effettivo della faglia di Brevard, una stretta fascia di rocce multideformate che segna il confine tra la provincia di Blue Ridge e la provincia di Piedmont. Segue una possibile interpretazione geologica dei profili. Le rocce cristalline di origine principalmente continentale sono rappresentate in grigio. Le rocce metamorfiche a pieghe vicino alla superficie sono invece rappresentate da linee in grigio ondulate. Gli strati sedimentari sono in colore chiaro. Infine le faglie desunte sono le linee con le frecce che ne indicano il senso del movimento.

chilometri a nord-ovest di Augusta (Georgia). La scoperta più rilevante è stata quella di uno strato, digradante verso sud-est, di superfici riflettenti che si estendono dalla Valley and Ridge sotto la Blue Ridge e l'Inner Piedmont fra quattro e 10 chilometri sotto la superficie.

Sebbene le riflessioni possano essere state prodotte da parecchi tipi e da parecchie configurazioni di rocce, i materiali riflettenti sono stati interpretati da parte nostra come strati sedimentari stratificati. Questa interpretazione si basa su quattro ordini di prove. Primo, le riflessioni sotto la Blue Ridge e l'Inner Piedmont possono essere seguite fino a unità analoghe, alle quali si possono correlare, della provincia di Valley and Ridge, di cui si sa che le rocce sono sedimentarie. In effetti alcune parti dei profili di riflessione sismica della Blue Ridge e del Piedmont sono del tutto simili al profilo di riflessione sismica della Valley and Ridge. Secondo, la presenza di rocce sedimentarie nelle finestre della Blue Ridge indica che ivi le rocce cristalline coprono materiale sedimentario. Terzo, la scoperta di rocce carbonatiche inconsuete nella zona di Brevard fa pensare che esse siano state asportate da sottostanti strati sedimentari da attività lungo i bordi della faglia. Infine, i risultati del COCORP assomigliano a quelli dei rilevamenti per riflessione sismica degli attuali margini continentali, che naturalmente consistono di materiale sedimentario.

H. Clark e i suoi colleghi del Virginia

In nove diagrammi a blocchi è compendiata la storia tettonica degli Appalachi meridionali. La crosta continentale è in colore grigio scuro, la crosta oceanica è in bianco, e le rocce sedimentarie della piattaforma continentale sono in colore chiaro. La superficie solida della Terra che si estende verso il basso per un centinaio di chilometri è costituita da una serie di zolle rigide interagenti che possono venire assorbite nel mantello in corrispondenza delle zone di subduzione: profonde fosse oceaniche dove due zolle entrano in collisione e una sprofonda sotto l'altra. La collisione di un continente o di un arco insulare oceanico con un complesso di zone di subduzione fa corrugare la roccia superficiale. Il corrugamento diede origine a tre orogenesi. Circa 650 milioni di anni fa, nel Precambriano *(1)*, un supercontinente si fratturò in due zolle continentali (Laurentia, o proto-America Settentrionale, e Gondwana, o proto-Africa e proto-America Meridionale) e in parecchi frammenti più piccoli, fra cui il Piedmont e la fascia ad argilloscisti della Carolina. La subduzione diede origine a vulcanismo nella fascia della Carolina probabilmente circa 625 milioni di anni fa *(2)*. L'attività vulcanica durò fino a circa 500 milioni di anni fa *(4)*. Forse l'orogenesi taconica, avvenuta da 500 a 450 milioni di anni fa, fu dovuta al chiudersi del bacino fra la proto-America Settentrionale e il Piedmont *(3, 4)*. Lo scorrimento di una faglia sottile che iniziò allora potrebbe avere due cause. La prima *(4, 5)* implica la subduzione continua verso est della parte inferiore del Piedmont mentre i 10 o 15 chilometri superiori si staccano. La seconda *(5')* implica la subduzione verso ovest della parte inferiore dell'Inner Piedmont e della Blue Ridge mentre la parte superiore incomincia a formare la faglia trascorrente. Il chiudersi del bacino oceanico tra l'Inner Piedmont e la fascia della Carolina *(5, 6)* provocò l'orogenesi acadiana, fra 400 e 350 milioni di anni fa, e il continuo scorrimento del Piedmont e del continente nordamericano. L'ultima orogenesi, quella allegheniana, mise la proto-Africa contro l'America Settentrionale *(6, 7)*. L'apertura dell'Atlantico, circa 200 milioni di anni fa, separò i continenti *(8)*.

Polytechnic Institute hanno fatto studi di riflessione sismica limitati alla zona situata 100 chilometri a nord di quella del rilevamento del COCORP nella North Carolina. Da questi studi risulta che anche sotto quella parte della Blue Ridge sono presenti strati sedimentari stratificati. Tali strati pertanto sostengono probabilmente buona parte degli Appalachi meridionali. Leonard Harris e Ken Bayer, del Geological Survey degli Stati Uniti, hanno analizzato dati ottenuti di recente per riflessione sismica raccolti nella North Carolina, i quali ancora una volta indicano la presenza di stratificazioni che si estendono dalla Valley and Ridge sotto la Blue Ridge fino al Piedmont.

L'unica spiegazione di questi strati sepolti è che le sovrastanti rocce cristalline siano state dislocate lungo un'importante faglia di compressione suborizzontale (ossia una faglia orizzontale sotto la superficie). I dati del COCORP indicano anche che quella di Brevard è una faglia di compressione che si estese, o si staccò, dalla più importante faglia di compressione suborizzontale. Non è da escludere che questo distacco abbia dato origine anche ad altre faglie della zona.

Lo studio del COCORP segue le tracce delle riflessioni stratificate in senso orizzontale fin sotto l'Inner Piedmont e la fascia di Charlotte. Poiché tali riflessioni sono simili e lateralmente contigue a quelle provenienti da sotto la Blue Ridge, è nostra opinione che dalla Valley and Ridge alla fascia di Charlotte gli strati sedimentari non abbiano soluzione di continuità. Ciò nonostante, il carattere delle riflessioni presenta un notevole cambiamento a una distanza di circa 250-300 chilometri dall'estremità nordoccidentale del profilo. Una serie di riflessioni digradanti verso est, simili a quelle che spesso caratterizzano i depositi su una scarpata continentale, contraddistingue questa parte della sezione sismica dalla banda di riflessione sottile e stratificata che va verso ovest.

Alla superficie, sopra le riflessioni digradanti, vi sono i graniti di Elberton, un grande corpo di roccia ignea intruso nella crosta. Sotto le rocce a pieghe e metamorfosate della parte orientale della fascia di Charlotte c'è un'altra ben distinta sequenza orizzontale e stratificata di riflessioni. Il cuneo delle riflessioni inclinate sia sotto i graniti di Elberton sia sotto la Charlotte, insieme con le riflessioni orizzontali provenienti dalla zona sottostante la stessa fascia, fa pensare che sotto la superficie, a una profondità compresa fra 12 e 18 chilometri, esiste una sequenza stratificata di rocce di grande spessore.

Altri mutamenti significativi nella natura delle onde riflesse dalle profondità della crosta si trovano all'estremità sudorientale della zona studiata, dove non è da escludere che riflessioni orizzontali anomale di 10,5-11 secondi (equivalenti a profondità comprese fra 30 e 33 chilometri) corrispondano alla transizione fra la crosta e il mantello. Questa transizione è nota come discontinuità di Mohorovičić. La discontinuità di Mohorovičić, così come altre superfici profonde, riflette onde diverse da quelle rilevate nei due terzi nordoccidentali del profilo. Altri dati sismici fanno pensare che sotto l'Inner Piedmont e la Blue Ridge la crosta possa scendere in profondità fino a 40-45 chilometri. Sopra la discontinuità di Mohorovičić, all'estremità meridionale, vi sono interessanti riflessioni digradanti verso ovest, che saranno difficili da interpretare finché il profilo non verrà esteso a est. Potrebbe trattarsi di zone di faglia, parti di un antico complesso di zone di subduzione, o forse addirittura di una struttura geologica di epoca anteriore non affine agli Appalachi.

Non è ancora dato sapere che cosa rappresentino i mutamenti delle riflessioni rilevati in corrispondenza della transizione fra l'Inner Piedmont e la fascia di Charlotte. Secondo una delle interpretazioni, le riflessioni sotto la fascia di Charlotte digradanti verso est segnano i punti in cui la più importante faglia di compressione si fa più ripida e scende in profondità nella crosta. Dato però che questa interpretazione non può spiegare facilmente delle riflessioni orizzontali a sud-est sotto la fascia ad argilloscisti della Carolina noi ne preferiamo una diversa. A nostro parere infatti le riflessioni orizzontali e la fitta sequenza di riflessioni digradanti verso est sotto la fascia di Charlotte sono causate da rocce sedimentarie che facevano parte di un margine continentale e del bordo di una piattaforma continentale ora sepolti sotto la faglia trascorrente. Oggi queste rocce sedimentarie sono molto profonde e quindi è probabile che si siano metamorfosate.

Questa interpretazione può spiegare anche la diminuzione di profondità della discontinuità di Mohorovičić da nord-ovest a sud-est lungo la zona d'indagine: è il risultato di una transizione da una crosta continentale a ovest a un'antica crosta oceanica o crosta continentale assottigliata a est. Questo lascia supporre che, al pari delle rocce deformate della Blue Ridge e dell'Inner Piedmont, le rocce deformate della fascia di Charlotte e della fascia ad argilloscisti della Carolina siano state spinte su un cumulo di sedimenti. Secondo questa interpretazione, è logico supporre che lo scorrimento continui più a est. Studi sismici attualmente in corso dovrebbero rivelare se questa ipotesi corrisponda o meno alla realtà.

Se le riflessioni sotto la fascia di Charlotte e quella ad argilloscisti della Carolina provengono effettivamente da strati sedimentari o sedimentari metamorfosati, prima che avesse avuto inizio lo scorrimento le rocce estremamente deformate o metamorfosate della Blue Ridge dovevano essere molto distanti dall'attuale posizione rispetto alla parte continentale interna. Probabilmente si trovavano a est di dove si trova ora la fascia ad argilloscisti della Carolina. La distanza di circa 260 chilometri che separa il bordo sudorientale dell'area di cui è stato tracciato il profilo sismico dal bordo occidentale della Blue Ridge è la distanza minima alla quale furono trasportate le rocce della Blue Ridge.

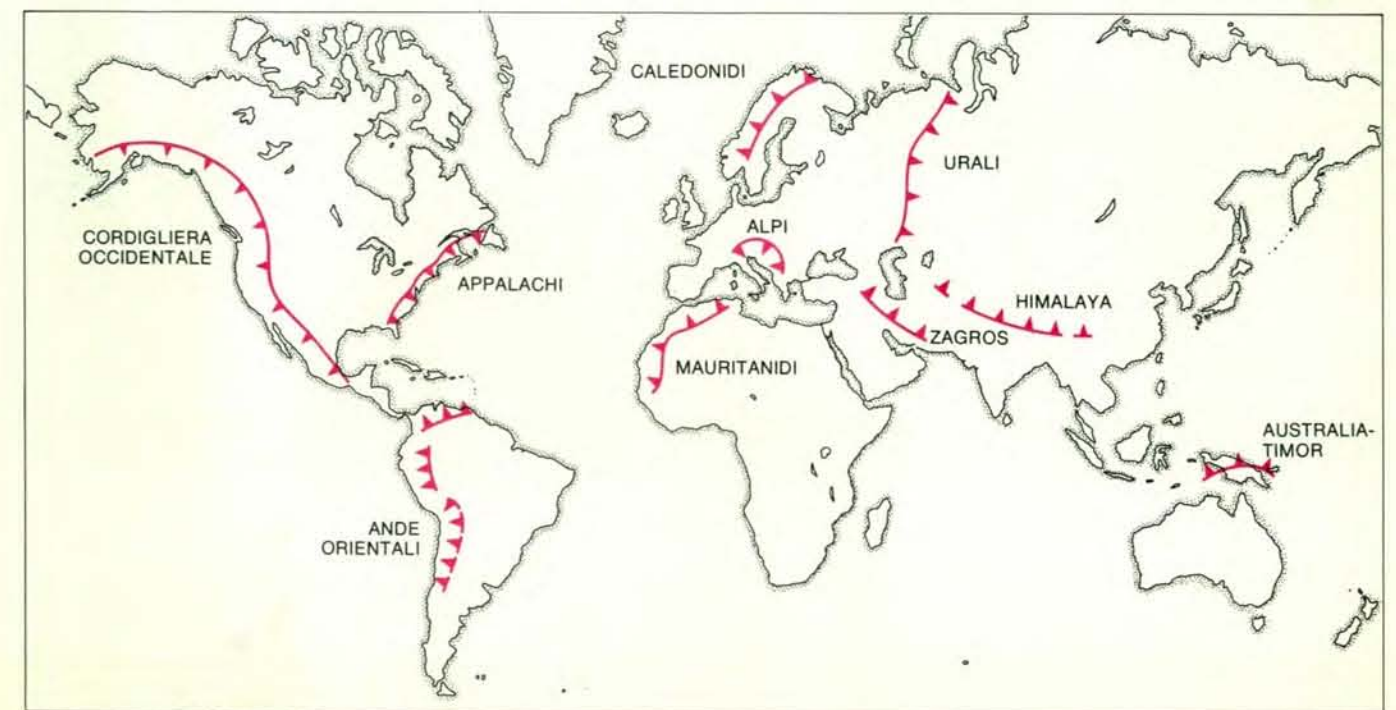
Che cosa rivela, della storia tettonica degli Appalachi, il profilo di riflessione sismica? Secondo la teoria della tettonica a zolle, le zolle rigide della litosfera, la superficie solida della Terra che si estende verso il centro per circa 100 chilometri, galleggiano sull'astenosfera, uno strato mobile e abbastanza fluido del mantello dello spessore di qualche centinaio di chilometri. Ogni zolla è costituita da uno strato superiore di crosta (il cui spessore varia da 35 a 40 chilometri per le zolle continentali e da 5 a 10 chilometri per quelle oceaniche) e da uno strato inferiore di mantello solido e compatto. La litosfera oceanica viene generata in corrispondenza delle dorsali medio-oceaniche, dove si ha una continua fuoriuscita di magma, che si raffredda e si indurisce in modo da formare il bordo trainante della zolla in movimento. Alla fine questa litosfera viene riassorbita nel mantello nelle zone di subduzione, le profonde fosse oceaniche che si vengono a creare dove due zolle entrano in collisione e una affonda sotto l'altra.

Una zona di subduzione continua a consumare litosfera finché un continente o un arco di isole oceaniche non urtano contro di essa. Poiché il materiale della crosta continentale è molto più leggero di quello del mantello, i geologi sono quasi tutti dell'avviso che difficilmente un continente possa essere subdotto come un tutto unico. Se la cosa non è effettivamente possibile, allora la collisione di un continente o di un arco insulare con un complesso di zone di subduzione muterà drasticamente la natura della subduzione o addirittura la impedirà. La deformazione della litosfera nel sito della collisione potrebbe dare origine a una catena montuosa.

La geologia del sottosuolo degli Stati Uniti orientali è stata desunta dai profili ottenuti per riflessione sismica dal COCORP. (Qui un tratto pianeggiante dell'area considerata.) Gli strati sedimentari sono in colore, le rocce cristalline di origine prevalentemente continentale in grigio scuro, e la crosta oceanica basaltica in bianco. Tra 500 e 250 milioni di anni fa la piattaforma continentale fu ricoperta da uno strato sottile quasi orizzontale di roccia cristallina. Una nuova piattaforma continentale si va formando al largo della costa atlantica da 200 milioni di anni.

Gli Appalachi meridionali si sono evoluti in una serie di collisioni di frammenti di materiale continentale o di archi insulari in corrispondenza del bordo orientale dell'America Settentrionale nelle orogenesi taconica, acadiana e alleghaniana. Sono stati elaborati parecchi modelli del processo evolutivo, e a questi modelli i dati del COCORP pongono nuovi vincoli di estrema importanza. Qui ci soffermeremo in particolare su un modello promettente, elaborato da Hatcher in modo da concordare con i dati geologici della superficie. Questo modello è stato da noi modificato tenendo conto delle informazioni ottenute dal COCORP sulla geologia al di sotto della superficie. I dati degli episodi orogenetici sono stati stabiliti sulla base della deposizione di sedimenti nel bacino appalachiano e della datazione radiometrica delle rocce ignee e metamorfiche. La composizione complessiva delle masse in collisione è stata desunta dal tipo delle rocce presenti nelle varie «province» e nelle varie «fasce».

Circa 750 milioni di anni fa magma risalente dall'interno della Terra divise una massa supercontinentale in almeno due grandi continenti (Laurentia, o proto-America Settentrionale, e Gondwana, o proto-Africa) e in almeno due frammenti continentali che comprendevano il frammento Inner Piedmont-Blue Ridge e quello della fascia ad argilloscisti della Carolina. Nel periodo iniziale di questo processo di fessurazione, le rocce che ora costituiscono il materiale vulcanico e sedimentario metamorfosato della Blue Ridge furono depositate nel bacino che separava la proto-America Settentrionale dal frammento Inner Piedmont-Blue Ridge.

Non esistono prove incontestabili che i frammenti del Piedmont e della fascia ad argilloscisti della Carolina provengano dallo stesso supercontinente che aveva formato l'America Settentrionale. Ciò nonostante, il fatto che alcune rocce della Blue Ridge e del Piedmont abbiano la stessa età (circa un miliardo di anni), stabilita col metodo della datazione radiometrica, delle rocce del basamento della parte orientale dell'America Settentrio-

Non è da escludere che le catene montuose di tutto il mondo (Cordigliera occidentale, Ande orientali, Mauritanidi, Alpi, Urali, Caledonidi, Zagros, Himalaya e Australia-Timor) siano costituite, come avviene nel caso degli Appalachi meridionali, da sottili faglie trascorrenti.

nale fa pensare che questi pezzi crostali siano stati sottoposti a metamorfismo nel Precambriano più o meno alla stessa epoca. Non è quindi da escludere che facessero parte dello stesso blocco continentale. La struttura degli Appalachi è il risultato non di un unico blocco crostale entrato in collisione con l'America Settentrionale, ma più probabilmente di una serie di collisioni di piccoli frammenti continentali o di archi insulari avvenute in un modo simile a quello delle numerose collisioni che avvengono oggi nel Pacifico sudoccidentale a mano a mano che l'Australia si muove verso l'Asia con un gruppo di archi insulari e di frammenti continentali intrappolati nello spazio intermedio.

Nell'arco insulare del frammento della fascia ad argilloscisti della Carolina il vulcanismo ebbe inizio circa 650 milioni di anni fa. Questo significa che anche la subduzione, che diede l'avvio all'attività vulcanica, ebbe inizio più o meno alla stessa epoca. Circa 500 milioni di anni fa il bacino fra la proto-America Settentrionale e il frammento Inner Piedmont-Blue Ridge incominciò a chiudersi in conseguenza della subduzione. L'esistenza stessa di strati sedimentari sotto la Blue Ridge e il Piedmont significa che la zona di subduzione affondò verso est, in quanto una subduzione verso ovest in quel momento avrebbe distrutto probabilmente i sedimenti quando si manifestarono i fenomeni di vulcanismo. Prima di 450-500 milioni di anni fa, i sedimenti che diedero origine alle arenarie e agli scisti che ora si trovano nella Valley and Ridge vennero dalla proto-America Settentrionale. Per i successivi 250 milioni di anni i sedimenti provennero principalmente dalla terraferma nella direzione opposta (ossia da oriente). Questo cambiamento nella storia sedimentaria coincide con il primo episodio di deformazione e di metamorfismo (l'orogenesi taconica, da 500 a 450 milioni di anni fa) e può essere attribuito al chiudersi del bacino tra il frammento Inner Piedmont-Blue Ridge e la proto-America Settentrionale e alla susseguente collisione di queste masse di terra. Le rocce sedimentarie dell'est venivano dalla faglia trascorrente che aveva incominciato a muoversi verso occidente in direzione del continente.

Non è per nulla chiaro come abbia fatto una faglia spessa dai 10 ai 20 chilometri a staccarsi dalla parte inferiore della crosta dell'Inner Piedmont e della Blue Ridge e a scorrere poi sopra la piattaforma continentale. Perché si spaccò, dove si spaccò, e che cosa ne fu dei rimanenti 80 chilometri o giù di lì della sottostante litosfera? È stata avanzata un'ipotesi, secondo la quale a mano a mano che la parte superiore della crosta si staccava, la restante parte inferiore della crosta e la parte superiore del mantello continuavano a essere subdotte verso est ed entravano in collisione con l'arco insulare della fascia ad argilloscisti della Carolina. Secondo un'altra ipotesi analoga, la polarità della subduzione nel bacino si invertì, di modo che una «scheggia» della parte superiore della crosta fu spinta verso il continente, mentre la parte inferiore della crosta dell'Inner Piedmont e della Blue Ridge veniva subdotta verso ovest. (L'ipotesi delle schegge era stato adottata in precedenza da E. R. Oxburg, dell'Università di Cambridge, per spiegare alcune caratteristiche delle Alpi.)

Il secondo episodio di sollevamento di una catena montuosa fu l'orogenesi acadiana, da 400 a 350 milioni di anni fa. Caratterizzata da deformazione e metamorfismo su vasta scala, questa orogenesi prese l'avvio dal chiudersi del bacino oceanico tra il frammento Inner Piedmont-Blue Ridge e il frammento della fascia ad argilloscisti della Carolina. A quel tempo probabilmente il frammento Inner Piedmont-Blue Ridge aderiva per sovrascorrimento alla proto-America Settentrionale. Non è da escludere che l'odierna fascia della Kings Mountain sia costituita dai resti superficiali dell'antica zona di collisione fra l'arco insulare della fascia ad argilloscisti della Carolina e l'Inner Piedmont. Anche se non si conoscono le dimensioni della fascia ad argilloscisti della Carolina, il frammento poteva benissimo essere molto grande, in quanto sotto la pianura costiera si estendono rocce metamorfiche simili alle rocce della fascia della Carolina.

Dopo l'orogenesi acadiana, il più importante (e ultimo) fenomeno di compressione negli Appalachi meridionali fu l'orogenesi allegheniana avvenuta nel periodo compreso tra 300 e 250 milioni di anni fa. Questo episodio di sollevamento di una catena montuosa può essere attribuito alla collisione fra la proto-America Settentrionale e la proto-Africa (o forse America Meridionale), in modo da formare il supercontinente Pangea. Sebbene negli Appalachi settentrionali l'orogenesi allegheniana non fosse così rilevante come le altre due orogenesi, negli Appalachi meridionali essa diede l'avvio a fenomeni di sovrascorrimento su vasta scala e a un'estesa attività eruttiva.

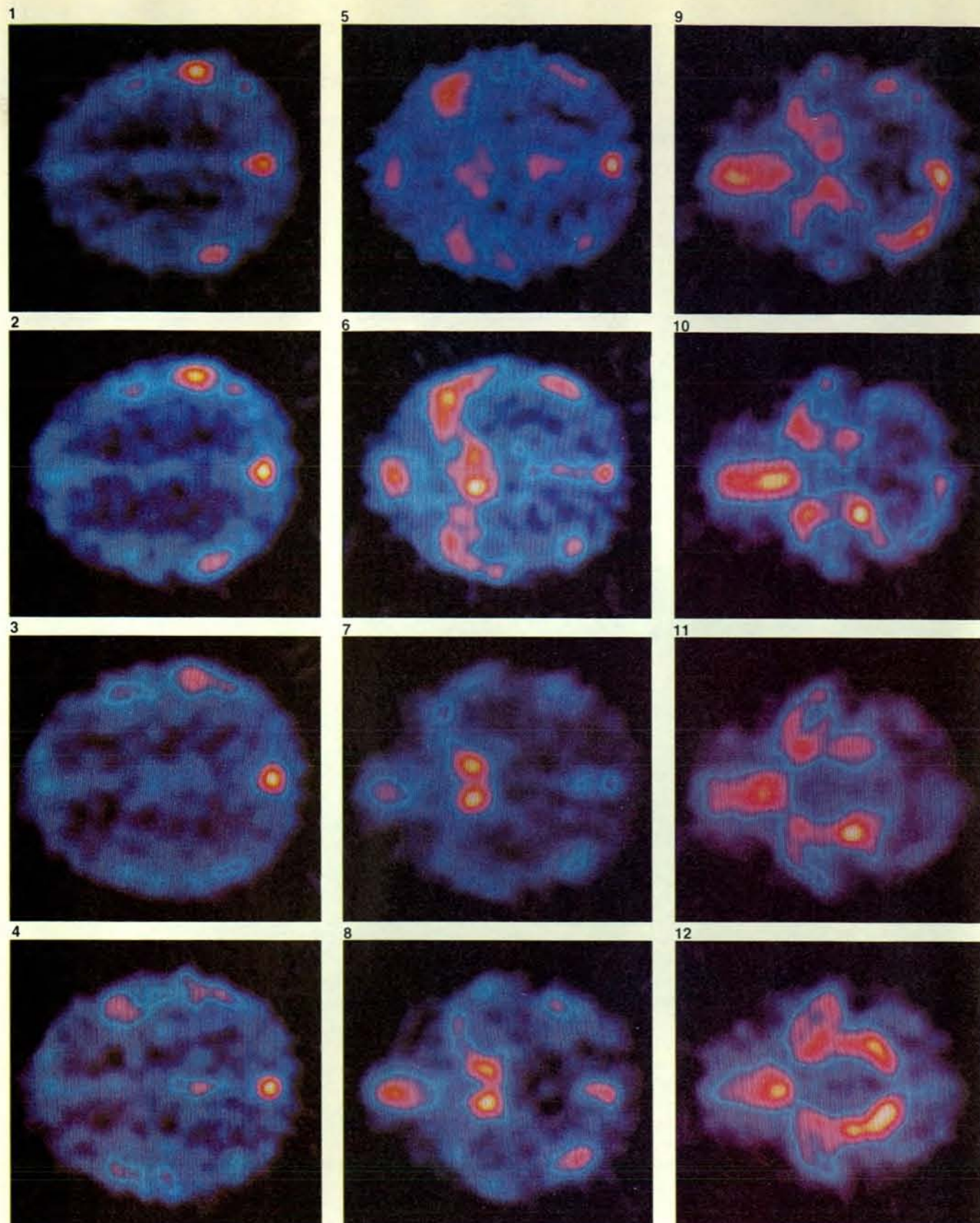
A quel tempo la zona di Brevard si aprì un varco fino alla superficie, portando con sé dal basso rocce sedimentarie carbonatiche. Dalla datazione radiometrica risulta che il dislocamento di molti corpi di rocce ignee del Piedmont ebbe luogo 300 - 250 milioni di anni fa. Un sovrascorrimento su vasta scala in questo stesso periodo è stato rilevato anche nell'Africa occidentale, anche se non si è ancora arrivati a precisare il limite occidentale delle faglie di compressione. Secondo noi, un segmento della piattaforma continentale africana (o sudamericana) si mosse sotto il margine orientale della fascia ad argilloscisti della Carolina, generando una fascia di pieghe e di faglie che andava nella direzione opposta.

Dopo l'orogenesi allegheniana, un'attività tettonica molto estesa incominciò ancora, da 250 a 200 milioni di anni fa, a dividere un supercontinente in continenti più piccoli. A mano a mano che i continenti si separavano, nella loro scia si apriva l'oceano Atlantico. I bacini triassici della costa orientale degli Stati Uniti, come i graben (avvallamenti con i lati quasi verticali) del New Jersey e del Connecticut, si formarono a quel tempo. A mano a mano che l'Atlantico si estendeva, al largo della costa orientale dell'America Settentrionale (e al largo della costa occidentale dell'Africa e di quella settentrionale dell'America Meridionale) prendeva forma l'attuale piattaforma continentale.

È possibile che sottili faglie trascorrenti su vasta scala siano il risultato generale di collisioni fra continenti? Vi sono parecchie indicazioni in questo senso. Abbiamo parlato delle fasce di pieghe e di faglie dell'Africa occidentale e della parte orientale dell'America Meridionale, che hanno una struttura geologica del tutto simile a quella delle fasce appalachiane. David Gee, del Servizio geologico svedese, ha rilevato importanti faglie orizzontali di compressione nelle montagne Caledonidi della Scandinavia, alcune delle quali danno l'impressione di essere state dislocate di centinaia di chilometri.

Il tipo di scorrimento di faglie sottili rilevato nella provincia di Valley and Ridge può essere visto anche nelle fasce di pieghe e faglie del Montana e della cordigliera di Alberta, che fa parte delle Montagne Rocciose. Quest'area sarà oggetto in futuro di studi del COCORP intesi a stabilire il limite occidentale dello scorrimento. Può darsi che lo scorrimento di faglie sottili sia stato prevalente anche in altre catene montuose, fra cui le Alpi, l'Himalaya e i monti Zagros del Medio Oriente. Secondo alcuni geologi, il tipo di sovrascorrimento evidente negli Appalachi sta avendo luogo attualmente nella regione di Timor nel Pacifico sudoccidentale, dove la piattaforma continentale australiana si è mossa verso nord-ovest sotto l'arco insulare di Timor per forse 150 chilometri.

L'intrappolamento di rocce sedimentarie sotto grandi faglie metamorfiche ha conseguenze di grande rilievo non solo per la crescita dei continenti ma anche per il recupero di gas e di petrolio. La scoperta di materiale sedimentario sotto le rocce metamorfiche delle province di Blue Ridge e di Piedmont impone un riesame della possibilità di estrarre idrocarburi da quella parte degli Appalachi, in quanto non è da escludere che il metamorfismo dei sedimenti non sia stato tanto intenso da rimuovere tutti i depositi di idrocarburi dalle rocce sotto la faglia trascorrente.

La scoperta dello strato di sedimenti sotto il nucleo degli Appalachi ha implicazioni che vanno molto al di là di questa particolare catena montuosa. Potrebbe darsi che il meccanismo dello scorrimento di faglie sottili che riguarda le rocce del basamento sia stato responsabile della formazione di catene montuose per tutta la durata dei processi della tettonica a zolle. In tal caso l'evoluzione dei continenti è caratterizzata non da aggiunte laterali lungo delimitazioni verticali, ma dall'accavallarsi e riassestarsi di faglie trascorrenti. Molto lavoro rimane ancora da fare, ma c'è la sensazione da parte nostra che almeno per metà della storia della Terra questo tipo di scorrimento abbia potuto benissimo costituire uno dei più importanti meccanismi dell'evoluzione e della crescita dei continenti.

Il volume di sangue nel cervello di un uomo normale è qui evidenziato in immagini che sono state realizzate a 12 livelli diversi per mezzo della tomografia a emissione di positoni (TEP). Il disegno nella pagina a fronte identifica i livelli corrispondenti alle immagini TEP qui numerate. Le regioni multicolori nelle immagini illustrano la distribuzione nella testa e nel cervello del tracciante radioattivo carbonio 11 (^{11}C), che ha un periodo di dimezzamento di soli 20 minuti. Il soggetto inala il radioisotopo nella forma di monossido di carbonio contenente ^{11}C, che si fissa all'emoglobina negli eritrociti del sangue e che viene trasportato rapidamente in tutto l'organismo. Quando un atomo di carbonio 11 decade, libera un positone, o elettrone positivo, ad alta energia. Il positone interagisce quasi istantaneamente con un elettrone e le due particelle si annichilano reciprocamente con emissione di due raggi gamma, che si propagano in direzioni quasi opposte. Le immagini TEP sono ricostruite al calcolatore a partire da misurazioni della radiazione. In queste immagini le regioni in cui è presente un grande volume di sangue appaiono in tonalità di rosso, giallo e bianco e delineano il grande sistema di raccolta del sangue venoso distribuito sulla superficie del cervello. Le tonalità di azzurro rappresentano generalmente il volume del sangue presente nei tessuti del cervello. Queste immagini TEP e le successive che illustrano questo articolo sono state eseguite dagli autori presso la School of Medicine della Washington University.

Tomografia a emissione di positoni

Questa tecnica rivela con maggiore chiarezza la concentrazione nei vari organi di sostanze marcate con isotopi radioattivi e promette grandi risultati tanto nella ricerca biologica quanto nella diagnosi clinica

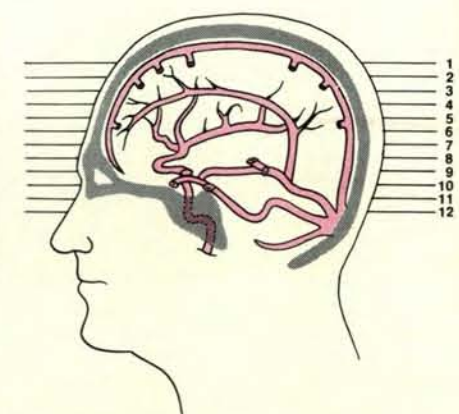
di Michel M. Ter-Pogossian, Marcus E. Raichle e Burton E. Sobel

Una premessa centrale della medicina è che ogni attività biologica è il risultato di reazioni biochimiche e che alla base di ogni fenomeno patologico vi è un difetto biochimico. Uno scopo costante della medicina è perciò quello di identificare l'attività biochimica anormale associata a una data patologia e di osservare tale anormalità direttamente e il più presto possibile nell'organismo malato. Una tecnica utile, sviluppata grazie alla disponibilità di isotopi radioattivi artificiali, consiste nel somministrare a un paziente una sostanza marcata con un radioisotopo e nel seguirne le vicende all'interno dell'organismo per mezzo di uno strumento in grado di rivelare il decadimento dell'isotopo. Le apparecchiature convenzionali per la rivelazione usate dalla medicina nucleare forniscono immagini un po' distorte della distribuzione di un isotopo radioattivo o tracciante in quanto il loro campo visuale varia in funzione della profondità e anche perché non riescono a differenziare la regione che interessa dalle regioni radioattive situate davanti o dietro di essa. Tali apparecchiature, in sostanza, comprimono tre dimensioni in due. A un certo numero di fini, tuttavia, le scintigrafie della medicina nucleare convenzionale sono del tutto idonee alle esigenze delle applicazioni cliniche.

Negli ultimissimi anni una localizzazione più precisa della radiazione emessa è stata resa possibile dallo sviluppo di una tecnica analoga a quella della tomografia assiale computerizzata (TAC), che costruisce un'immagine facendo ruotare una sorgente di raggi X attorno a un soggetto umano. Un rivelatore, situato in direzione diametralmente opposta alla sorgente, ruota di moto sincrono con essa attorno al soggetto in modo da registrare le variazioni subite dalla radiazione nel suo passaggio attraverso i tessuti interposti. Un calcolatore ricostruisce l'immagine nella forma di una sezione assiale trasversale. È evidente che quando una condizione patologica si manifesta o in un'immagine convenzionale ai raggi X o in un'immagine TAC, la anormalità biochimica che ne è causa è di solito già molto avanzata.

La tecnica che combina la valutazione biochimica precoce della patologia conseguita dalla medicina nucleare con la precisa localizzazione consentita dalla ricostruzione computerizzata dell'immagine è la tomografia a emissione di positoni (TEP). In questa tecnica un composto chimico che possiede l'attività biologica desiderata viene marcato con un isotopo radioattivo che decade emettendo un positone (ossia un elettrone con carica positiva). Il positone emesso si combina quasi immediatamente con un elettrone e le due particelle si annichilano reciprocamente con emissione di due raggi gamma. Questi si propagano rapidamente in direzioni quasi esattamente opposte, penetrano nel tessuto circostante e vengono registrati all'esterno del soggetto da una serie di rivelatori disposti in cerchio. Un algoritmo matematico applicato dal calcolatore ricostruisce rapidamente la distribuzione spaziale della radioattività all'interno del soggetto per un piano scelto e presenta l'immagine che ne risulta su uno schermo a raggi catodici. Dopo la somministrazione del composto marcato, le immagini vengono registrate a intervalli e possono essere codificate in colore allo scopo di evidenziare le differenze nel livello di attività da un punto all'altro. Le immagini TEP, opportunamente interpretate, possono fornire una stima localizzata non invasiva di molti processi biochimici che sono essenziali al funzionamento dell'organo esaminato.

Finora la TEP è stata molto usata su animali da esperimento, primariamente su cani e scimmie, e in misura più limitata su oggetti umani. Essa ha già permesso di gettare nuova luce su molti processi fisiologici che non possono essere studiati in modo altrettanto efficace con altre tecniche. Ancora pochissimi anni or sono la TEP era considerata una tecnica esotica e costosa, limitata a pochi laboratori. Oggi sta per diventare uno strumento di fondamentale importanza nella medicina diagnostica ed è già usata in una quarantina di importanti centri medici in tutto il mondo.

Oltre a fornire localizzazioni precise della sostanza marcata somministrata al soggetto, la TEP ha un secondo vantaggio importante rispetto alle tecniche convenzionali della medicina nucleare. Essa offre infatti una scelta più utile di elementi chimici biologicamente significativi per la marcatura. I sistemi biologici constano principalmente di composti del carbonio, dell'azoto, dell'ossigeno e dell'idrogeno. I primi tre elementi hanno isotopi radioattivi di breve vita che decadono con emissione di positoni e perciò sono idonei per la tomografia a emissione di positoni. Per gli studi TEP tre isotopi di particolare valore sono l'ossigeno 15, l'azoto 13 e il carbonio 11, che hanno rispettivamente

I piani delle immagini TEP che compaiono nella pagina a fronte sono identificati in questa sezione trasversale. Le principali strutture venose del cervello, che spiccano nelle immagini TEP, sono rappresentate qui in colore.

periodi di dimezzamento di circa due, 10 e 20 minuti. Benché non esista alcun isotopo dell'idrogeno che emetta positoni (il trizio, o idrogeno 3, emette solo elettroni a bassa energia), l'acqua può essere marcata per studi con la TEP usando ossigeno 15. Prima dello sviluppo della TEP, il carbonio 11, l'ossigeno 15 e l'azoto 13 avevano conosciuto una limitata applicazione in studi di biochimica in connessione con rivelatori convenzionali per registrare i raggi gamma creati dall'annichilazione di positoni.

Lo sviluppo della TEP fu ritardato in grande misura dalla brevità dei periodi di dimezzamento degli isotopi degli elementi traccianti più utili. Poiché la loro esistenza è così effimera, il carbonio 11, l'ossigeno 15 e l'azoto 13 devono essere prodotti molto vicino al punto in cui devono essere rivelati e tecniche chimiche rapide devono quindi essere escogitate per introdurre gli isotopi indicatori in composti utili. Infine, gli studi TEP con i composti marcati sono limitati a processi biochimici con tassi di ricambio rapidi.

Nel corso degli ultimi vent'anni le difficoltà associate alla produzione e alla manipolazione di tali isotopi di breve vita sono state in gran parte risolte. Oggi vari centri sono attrezzati con ciclotroni progettati espressamente per creare una varietà di isotopi che emettano positoni. I ciclotroni che hanno questo scopo accelerano tipicamente protoni (nuclei di idrogeno) sino a far loro raggiungere un'energia di 15 milioni di elettronvolt (MeV) e deutoni (nuclei di idrogeno pesante o deuterio) sino a far loro raggiungere la metà di tale energia. Un ciclotrone destinato a questi usi pesa una ventina di tonnellate ed è schermato da spessi muri di cemento.

I materiali-bersaglio che vengono sottoposti a bombardamento nel ciclotrone variano a seconda dell'isotopo radioattivo desiderato e della forma chimica in cui esso emerge dal bersaglio. Spesso il bersaglio è un gas e questa scelta è dovuta al desiderio di facilitare il trasporto del materiale attivo dal ciclotrone all'area in cui dev'essere usato. Anche quando il bersaglio è un solido, l'isotopo così formato viene liberato abitualmente in forma gassosa. Il carbonio 11 viene preparato di solito bombardando azoto o boro con protoni, l'ossigeno 15 irradiando azoto con deutoni e l'azoto 13 irradiando ossigeno con protoni o carbonio con deutoni. Le reazioni vengono scelte in modo che i radioisotopi vengano sintetizzati in forma relativamente esente da contaminazione a partire da isotopi stabili dello stesso elemento. Tali isotopi sono molto desiderabili per vari procedimenti di marcatura.

In alcuni casi l'isotopo radioattivo può essere usato direttamente nella sua forma elementare. Così l'ossigeno 15 viene utilizzato in studi sul metabolismo e l'azoto 13 in studi sulla ventilazione dei polmoni. Un certo numero di composti semplici vengono marcati facilmente: l'acqua con ossigeno 15, il monossido di carbonio o l'anidride carbonica con carbonio 11 o con ossigeno 15 e l'ammoniaca con azoto 13. In altri casi il problema di marcare un composto complesso può dar filo da torcere per mesi o anni al chimico più ricco di immaginazione prima che egli riesca a escogitare un procedimento di marcatura abbastanza rapido da risultare compatibile col periodo di dimezzamento dell'isotopo. Per esempio, la marcatura di glucosio con carbonio 11 si rivelò notevolmente difficile, benché quella con carbonio 14 (che ha un periodo di dimezzamento di 5730 anni) fosse ben affermata. Foglie di barbabietola illuminate vengono esposte ad anidride carbonica marcata con carbonio 11. Attraverso la fotosintesi l'anidride carbonica viene incorporata nel glucosio, il quale viene poi estratto con metodi chimici rapidi e purificato per mezzo della cromatografia.

Talvolta ci si imbatte in difficoltà anche dopo avere escogitato un procedimento di marcatura rapido. Per esempio, la molecola marcata potrebbe non essere facilmente solubile nei liquidi organici. Questo problema si presentò con il palmitato, il sale dell'acido palmitico, un acido grasso usato nello studio dell'integrità del muscolo cardiaco. Il sale fu reso solubile legandolo alla proteina sieroalbumina, la proteina del siero del sangue. A volte il problema risiede nel conseguire un'attività estremamente specifica. Per esempio, i composti idonei per lo studio dei recettori di cellule nervose devono avere un'attività specifica di migliaia di curie per millimole. Ciò significa che devono essere marcate la maggior parte o tutte le molecole presenti in un dato campione. Nonostante le varie difficoltà, sono stati oggi elaborati metodi per marcare varie centinaia di specie di molecole con isotopi che emettono positoni, superando in tal modo il numero di molecole usate nella medicina nucleare convenzionale.

Un isotopo di particolare utilità negli studi TEP è il fluoro 18, che ha un periodo di dimezzamento di 110 minuti. Pur avendo l'ovvia proprietà di chiarire il ruolo del fluoro in traccia nella fisiologia animale, il fluoro 18 ha una virtù addizionale: esso può essere incorporato in molecole che sono analoghi di substrati naturali del metabolismo. All'interno dell'organismo, l'analogo segue il percorso metabolico della molecola normale sino a un certo punto, in cui la sequenza di reazione si arresta bruscamente in quanto la molecola analoga non è più accettata dal meccanismo biochimico della cellula. Un analogo del glucosio in cui il fluoro 18 è sostituito a uno degli atomi di ossigeno si è rivelato utile nello studio del metabolismo del glucosio.

La tomografia assiale computerizzata, o TAC, si fonda su programmi simili a quelli usati dalla TEP. In un'immagine TAC l'attenuazione dei raggi X per opera del tessuto lungo diversi assi è registrata man mano che la sorgente di radiazione viene spostata in cerchio attorno al soggetto. In un procedimento tipico, la sorgente di raggi X compie una traslazione in ciascuna delle 180 posizioni che sono separate fra loro di un grado. Le misurazioni eseguite in questo modo sono sufficienti a un calcolatore per ricostruire l'opacità relativa del tessuto in una sezione bidimensionale.

Il processo di formazione dell'immagine nella tomografia a emissione di positoni presenta tre fasi principali: la rivelazione dei raggi gamma emessi nel processo di annichilazione dei positoni, l'accertamento della direzione di propagazione delle radiazioni e la ricostruzione della distribuzione della radiazione in una precisa immagine geometrica. Un'immagine tomografica può essere ricostruita con l'aiuto di isotopi che emettono direttamente raggi gamma nel corso del loro decadimento, ma il numero di isotopi adatti è limitato. Inoltre, se l'isotopo decade attraverso l'emissione di positoni, la ricostruzione viene semplificata e l'immagine rappresenta più fedelmente la distribuzione del composto marcato. Il positone emesso perde la maggior parte della sua energia cinetica dopo aver percorso soli pochi millimetri nel tessuto vivente. A questo punto è molto soggetto a interagire con un elettrone, un evento che causa l'annichilazione di entrambe le particelle. La massa delle due particelle viene convertita in 1,02 milioni di elettronvolt di energia, divisa in parti uguali fra due raggi gamma, o fotoni ad alta energia.

Poiché i due raggi gamma vengono emessi simultaneamente e si propagano in direzioni quasi esattamente opposte, è possibile localizzarne la sorgente con grande precisione. Questo risultato viene ottenuto accoppiando i rivelatori di radiazione e situando ciascun rivelatore di ogni coppia a 180 gradi di distanza dall'altro; i due rivelatori registrano un segnale solo se entrambi percepiscono simultaneamente fotoni ad alta energia. In tal modo una serie di rivelatori disposti su una circonferenza possono individuare le sorgenti di tutte le coppie simultanee di raggi gamma che hanno origine all'interno di un volume definito dalle linee rette che uniscono i rivelatori accoppiati. Gli eventi di annichilazione che hanno luogo all'esterno del volume definito non vengono rivelati. La risoluzione spaziale conseguibile dipende dall'apertura dei rivelatori. La risoluzione è leggermente inferiore nel sito del decadimento radioattivo, in conseguenza del raggio d'azione dei positoni nella materia e del fatto che i fotoni risultanti dall'annichilazione delle coppie particella-antiparticella non sono emessi esattamente in direzioni a 180 gradi l'una dall'altra.

Il carattere distintivo di un'immagine TEP consiste nel fatto che la radiazione penetrante ha origine all'interno del soggetto anziché provenire dall'esterno. La sorgente della radiazione è trasportata nel tessuto da un composto biologicamente attivo, marcato con un isotopo che decade emettendo un positone che si annichila con un elettrone entro pochi millimetri dal sito del decadimento. I rivelatori sono disposti in modo da registrare solo eventi simultanei a 180 gradi di distanza fra loro. Un programma di calcolatore ricostruisce la distribuzione degli isotopi che decadono.

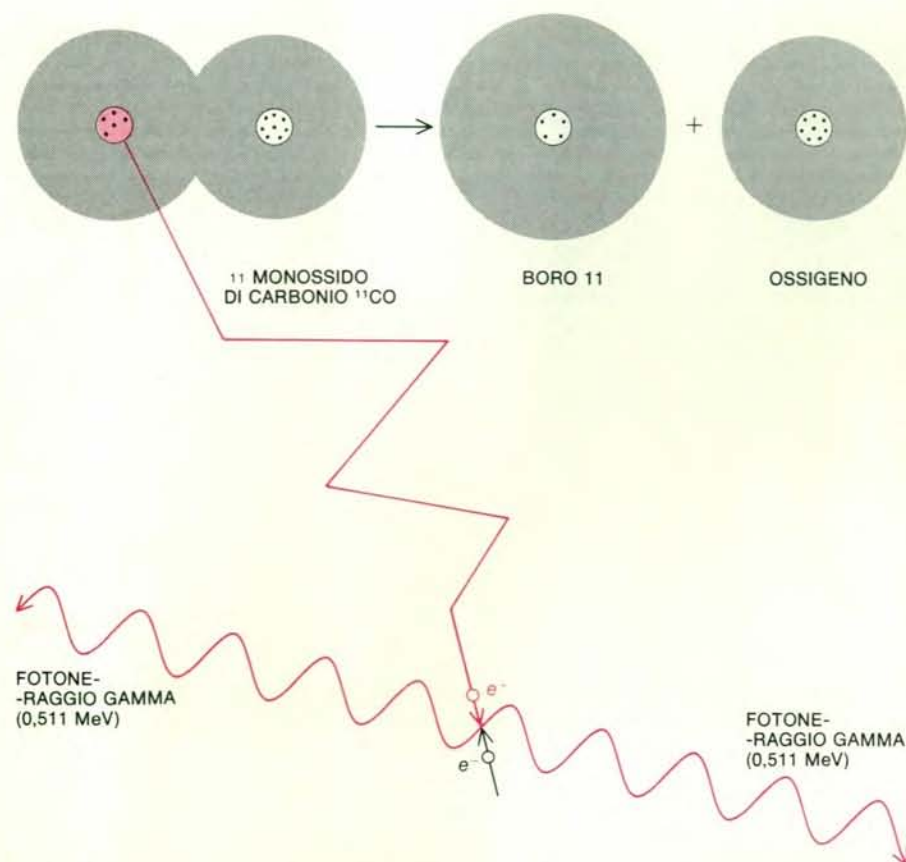
Gli eventi registrati dai rivelatori di coincidenza sono i dati grezzi a partire dai quali vengono ricostruite le immagini TEP. Il principio della ricostruzione dell'immagine nella tomografia assiale computerizzata è che un oggetto può essere riprodotto con precisione a partire da una serie di sue proiezioni prese ad angoli diversi. La fedeltà di una tale ricostruzione è proporzionale al numero di proiezioni. Dato un insieme di proiezioni infinito, un oggetto può essere ricostruito con precisione assoluta. In un sistema TEP tipico, un numero di proiezioni compreso fra 100 e 300 fornirà una risoluzione spaziale di pochi millimetri.

Misurando fotoni coincidenti, la serie di rivelatori in un sistema TEP identifica una serie di linee rette, situate tutte in un singolo piano. Quest'informazione viene selezionata in una serie di linee parallele fittamente accostate l'una all'altra, le quali rappresentano la radioattività in un soggetto osservato da angoli diversi. I gruppi di linee parallele, registrate sotto forma di profili, vengono corretti per mezzo di un processo matematico chiamato convoluzione, che attribuisce ai profili valori negativi destinati ad abolire le strutture artificiali presenti nell'immagine. I profili convoluti sono «proiettati all'indietro» per ottenere l'immagine finale dell'oggetto. Altri algoritmi che non utilizzano la convoluzione sono stati applicati con successo allo stesso scopo, ma la tecnica della convoluzione, o «proiezione all'indietro filtrata», è oggi quella di uso più generale.

Un'unità TEP completa consta di un sistema per l'acquisizione di dati e di un calcolatore. Il sistema di acquisizione è formato dai rivelatori di radiazione, dai circuiti a essi associati e, nella maggior parte dei tipi, da un sistema meccanico che impartisce un piccolo movimento ai rivelatori allo scopo di ottenere un migliore campionamento. In alcuni tipi alternativi il numero dei rivelatori è abbastanza grande da rendere superfluo il movimento. I dati forniti dal sistema di acquisizione vengono introdotti in un calcolatore veloce, di solito dotato di una grande memoria. Fra gli equipaggiamenti periferici del calcolatore vi sono dischi magnetici e nastri, una stampante parallela che fornisce valori numerici della distribuzione della radioattività nell'immagine, un sistema di visualizzazione per la osservazione diretta e la registrazione dell'immagine e spesso alcuni dispositivi di interazione per l'analisi dell'immagine. Alcuni sistemi TEP includono circuiti elettronici capaci di eseguire parti del processo di ricostruzione, che altri sistemi affidano invece per intero al calcolatore.

I sistemi TEP più recenti sono in grado di registrare simultaneamente sino a sette immagini tomografiche della testa o del tronco umani. Essi hanno una risoluzione di un centimetro circa e stimano la distribuzione della radioattività con un margine di errore di poche unità percentuali. Immagini successive possono essere ottenute a intervalli di meno di un minuto. La dose di materiale radioattivo somministrata al soggetto produce un'esposizione a radiazioni paragonabile a esposizioni tipiche in altri tipi di medicina diagnostica nucleare. Varie ditte produttrici offrono già sistemi TEP a prezzi che variano fra 600 000 e più di un milione di dollari.

Passiamo ora a considerare i tipi di studi fisiologici che possono essere eseguiti con la tomografia a emissione di positoni. Una fra le prime misurazioni compiute con la TEP fu la determinazione del volume regionale di sangue presente in sezioni trasversali del cervello umano. I risultati più soddisfacenti vengono ottenuti facendo inalare al soggetto aria contenente tracce di monossido di carbonio marcato con carbonio 11. Il monossido di carbonio contenente carbonio 11 si lega avi-

Isotopi che emettono positoni vengono creati bombardando elementi non radioattivi con particelle come protoni, deutoni e nuclei di elio. Per esempio, il carbonio 11 si produce quando il boro 10, il cui nucleo è formato da cinque neutroni e cinque protoni, cattura un altro protone *(a)*. In molti studi il carbonio 11 si combina con ossigeno per dare monossido di carbonio radioattivo *(b)*.

La sorte di un composto marcato all'interno di un organismo si evidenzia quando il radioisotopo incorporato come tracciante decade. Per esempio, il monossido di carbonio con ^{11}C funge da tracciante radioattivo per l'emoglobina formando ^{11}C-carbossiemoglobina. Quando un nucleo di carbonio 11 decade libera un positone e diventa boro 11, il cui nucleo è formato da cinque protoni e sei neutroni. Di conseguenza il monossido di carbonio viene dissociato in un atomo di boro e uno di ossigeno. Quando il positone (e^+) incontra un elettrone (e^-), ciascuno dei due raggi gamma risultanti dall'annichilazione trasporta poco più di mezzo milione di elettronvolt di energia.

damente con l'emoglobina, formando ^{11}C-carbossiemoglobina, e marcando in tal modo efficacemente l'intera quantità di sangue presente nell'organismo. Il volume del sangue presente nel cervello viene determinato rapidamente comparando la quantità di ^{11}C-carbossiemoglobina che si osserva nelle immagini tomografiche del cervello con la quantità presente nell'intera massa sanguigna quale è determinata sulla base di campioni di sangue venoso prelevati simultaneamente alla produzione dell'immagine tomografica. Misurazioni del volume di sangue nelle regioni cerebrali per mezzo della TEP non solo rivelano le differenze regionali attese, dovute al fatto che i vasi sanguigni nella materia grigia della corteccia cerebrale sono molto più fitti che nella sottostante materia bianca, ma consentono anche di delineare i vasi sanguigni principali, primariamente vene, che circondano il cervello.

Pur esistendo varie altre tecniche ben sperimentate per la misurazione del volume di sangue presente nel cervello, il successo del metodo TEP fu importante per varie ragioni. Innanzitutto dimostrò che la perfusione regionale di un organo come il cervello può essere stimata con un metodo semplice e preciso. In secondo luogo le tecniche TEP per la misurazione del volume di sangue cerebrale possono essere estese ad altri organi e tessuti marcando sostanze diverse dal monossido di carbonio. Per esempio, marcando piastrine del sangue è oggi possibile, per la prima volta, studiare il processo dell'aterosclerosi nei vasi sanguigni di un animale vivente. Una tale informazione dovrebbe essere di grande aiuto nella diagnosi e nel trattamento di pazienti umani affetti da varie forme di malattie vascolari, fra cui i blocchi delle arterie che causano colpi apoplettici e attacchi cardiaci.

In terzo luogo, la tecnica TEP per misurare il volume della massa sanguigna può essere adattata per determinare la concentrazione di qualsiasi composto marcato in modo idoneo, compresi farmaci, nel sistema vascolare di un organo-bersaglio, rispetto alla sua concentrazione nel compartimento extravascolare. La capacità di stabilire questa distinzione è esclusiva della TEP ed è molto importante nella misurazione del metabolismo, della composizione chimica dei tessuti e del flusso sanguigno.

La misurazione del metabolismo regionale è un po' più complessa della misurazione del volume regionale del sangue. Studi quantitativi del metabolismo cerebrale sono stati compiuti recentemente con ossigeno marcato con ossigeno 15, con glucosio marcato con carbonio 11 e con l'analogo del glucosio contenente fluoro 18 (^{18}F-2-desossi-D-glucosio). Anche il metabolismo del muscolo cardiaco è stato studiato utilizzando palmitato, il sale dell'acido palmitico, e substrato fisiologico primario del cuore, marcato con carbonio 11.

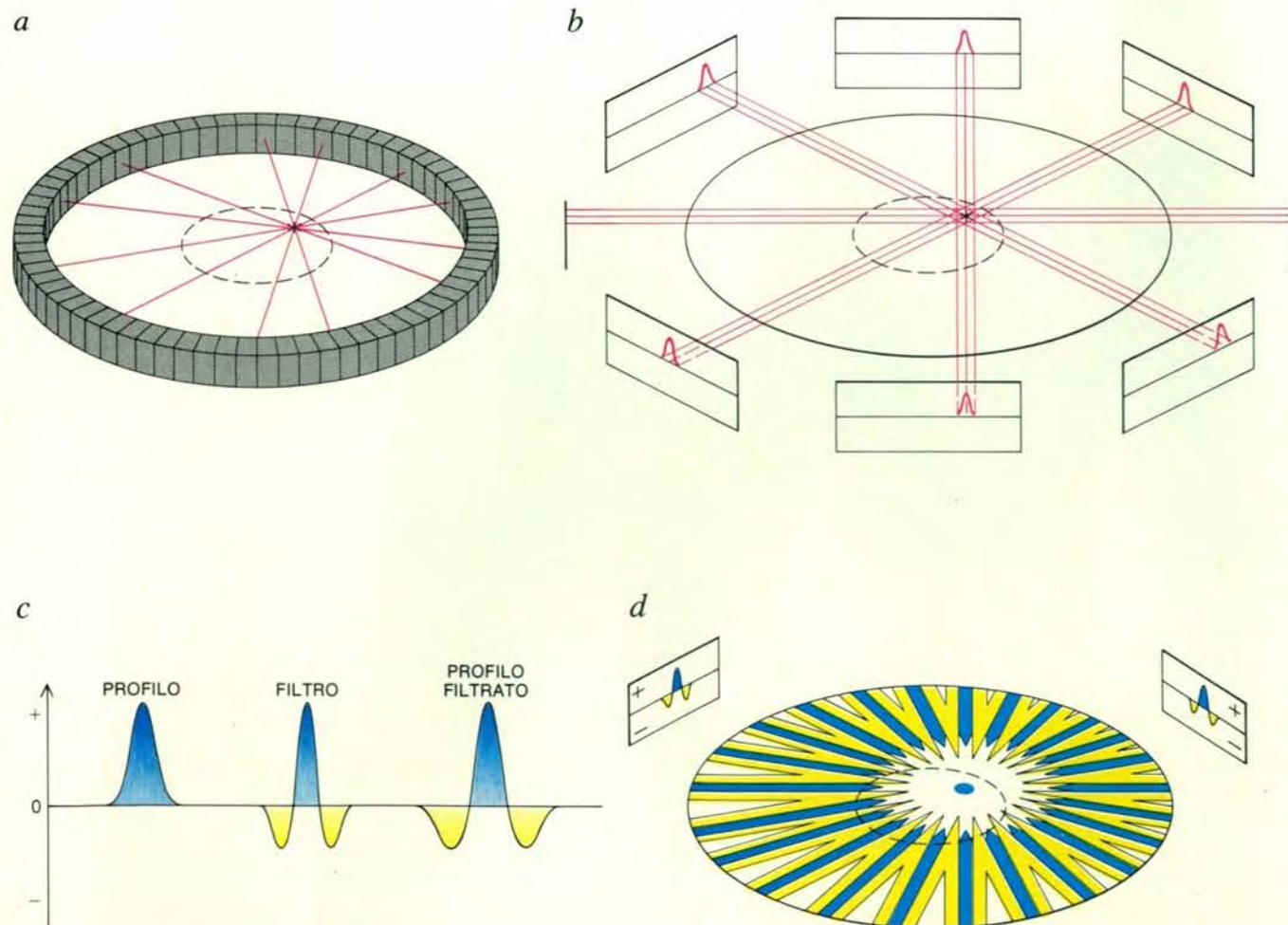
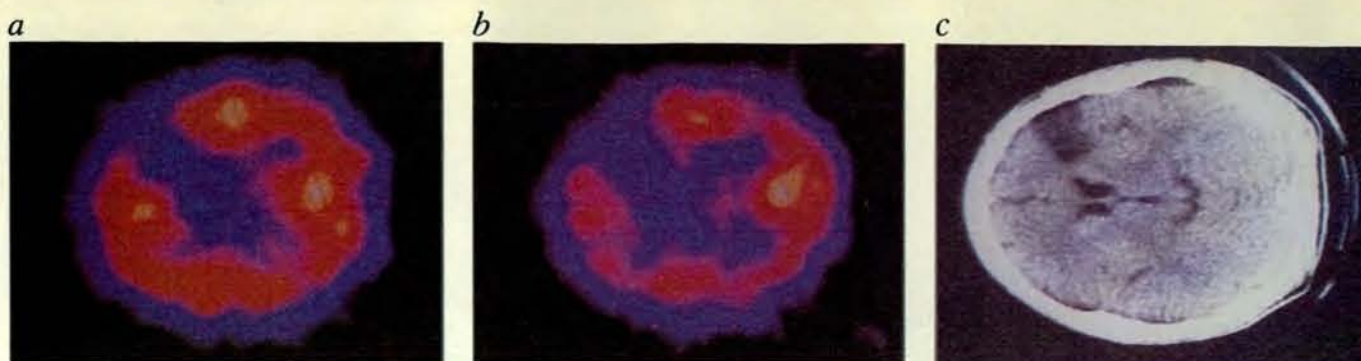
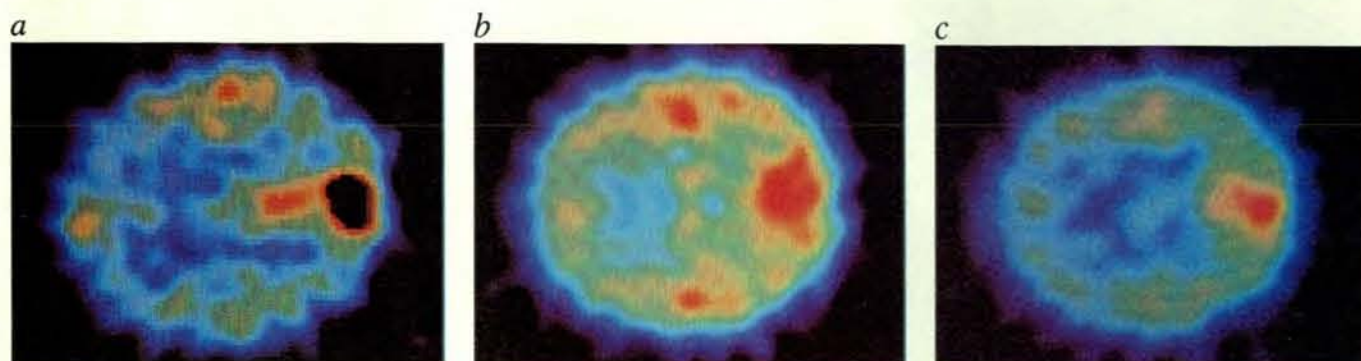
Recentemente sono state sviluppate varie tecniche promettenti per la diagnosi non invasiva in cardiologia. Esse

comprendono la stima con ultrasuoni (ecocardiografia) delle dimensioni dei ventricoli del cuore e del moto delle valvole, la valutazione della funzione di pompaggio del cuore per mezzo di radioisotopi somministrati per endovena (radioventricolografia) e la definizione dell'anatomia del cuore mediante immagini TAC a scansione. Questi metodi, così come anche altri metodi disponibili, non hanno i caratteri intrinseci necessari per una quantificazione definitiva della perfusione regionale e del metabolismo del muscolo cardiaco. Il problema cardiaco più comune in cui ci si imbatte negli adulti è l'ischemia: una condizione causata da un insufficente afflusso di sangue in regioni del cuore. I sintomi vanno da dolori episodici al torace (angina pectoris) a insufficienza circolatoria (fenomeni congestizi) e ancora a infarto del miocardio («attacco di cuore»).

L'insufficienza delle coronarie, che è la causa di queste affezioni, conduce non solo a sintomi, ma anche ad alterazioni del metabolismo regionale all'interno del cuore. Per mezzo delle tecniche TEP, traccianti come il carbonio 11 possono essere localizzati con precisione all'interno del muscolo cardiaco. Inoltre, quando i traccianti che emettono positoni sono incorporati in sostanze appropriate, come il palmitato, offrono la possibilità di caratterizzare il metabolismo regionale con una sostanza il cui comportamento chimico è identico a quello della sostanza fisiologica in studio.

Per poter definire la natura rivelabile all'esterno degli importanti fenomeni biochimici in atto nel cuore che possono essere stimati dalla TEP, sono stati compiuti studi su cuori isolati asportati da animali da esperimento e anche sul cuore di soggetti umani viventi. L'accumulazione di un qualsiasi tracciante in una regione del cuore dipende da vari fattori, fra cui il tasso di liberazione del tracciante nel tessuto (una funzione della perfusione regionale), il tempo di permanenza del tracciante, la frazione di esso che viene estratta a ogni singolo passaggio attraverso la circolazione coronarica e i tassi del metabolismo e dell'eliminazione del tracciante inizialmente estratto dal sangue per opera del muscolo cardiaco. La frazione di estrazione e il metabolismo dipendono in modo critico non solo dalla natura della sostanza marcata, ma anche dalle condizioni del metabolismo regionale del muscolo cardiaco al momento della sua valutazione.

Questo rapporto critico fu stabilito in studi su cuori isolati da animali e perfusi con un liquido nutritivo. A pochi secondi di intervallo dall'iniezione nel liquido perfuso di palmitato marcato con carbonio 11, l'immagine TEP registra un picco di massima attività che è proporzionale alla quantità del tracciante iniettato. L'attività massima declina poi rapidamente quando una frazione consistente del tracciante viene trasportata via dalla circolazione coronarica. Questo declino è seguito da uno più lento che rappresenta primariamente lo scambio e l'eliminazione graduale del tracciante dal liquido interstiziale. La terza e ultima fase del declino di attività segue una lenta curva esponenziale che riflette principalmente la sorte del palmitato marcato con ^{11}C incorporato in lipidi all'interno delle cellule del muscolo cardiaco.

Se si confronta questa fase finale con l'iniziale ascesa della curva della radioat-

Il processo TEP presenta quattro fasi principali. Nella prima fase *(a)* i fotoni dei raggi gamma che marcano i siti di decadimento di un composto che emette positoni vengono registrati da una serie circolare di rivelatori. Le misurazioni rappresentano una serie di linee rette, che si trovano tutte su un singolo piano. L'informazione viene scomposta in una serie di linee parallele *(b)* che rappresentano una proiezione ad angoli diversi della radioattività, all'interno di una regione prescelta del soggetto. Queste proiezioni, che vengono registrate inizialmente come profili, vengono passate matematicamente attraverso un filtro, in un procedimento chiamato convoluzione che fornisce un mezzo per eliminare dall'immagine strutture artificiali *(c)*. I profili convoluti vengono poi «proiettati all'indietro» per ricostruire un'immagine di come il composto marcato era distribuito *(d)*. La manipolazione di grandi quantità di dati raccolti richiede l'adozione di un calcolatore veloce.

Immagini di un infarto cerebrale dovuto a un colpo apoplettico ottenute con la TEP *(a* e *b)* e con la TAC *(c)*. Nelle immagini TEP tali infarti si evidenziano immediatamente, mentre un'immagine TAC rivela solamente lenti mutamenti nella densità del tessuto. La prima immagine TEP *(a)*, eseguita dopo che il paziente aveva ricevuto un campione di sangue marcato con acqua contenente ^{15}O, illustra il flusso cerebrale del sangue. La seconda immagine TEP *(b)* illustra il consumo di ossigeno dopo che il paziente aveva inalato ossigeno contenente ^{15}O.

Il sito di un tumore cerebrale, viene identificato in queste immagini TEP, relative al volume del sangue *(a)*, al flusso sanguigno *(b)* e al consumo d'ossigeno *(c)*. I meningiomi sono tumori molto vascolarizzati che si sviluppano nella membrana che ricopre il cervello. Questo è rivelato dall'accresciuto volume di sangue al centro, oltre a un aumento locale nel flusso sanguigno, e a una diminuzione nel consumo di ossigeno.

Il sito di un attacco focale in un epilettico è evidenziato in questa TEP, che presenta il volume del sangue nel cervello *(a)*, il flusso sanguigno *(b)* e il consumo di ossigeno *(c)*. Nell'immagine *a* il puntino rosso che appare in prossimità delle ore 12 si trova nel sito dell'attacco focale.

Un cuore normale e due danneggiati presentano vistose differenze in queste immagini TEP, che danno la risposta del cuore a due composti marcati con radioisotopi. Le aree verdi illustrano la distribuzione del carbonio 11 iniettato come ^{11}C-palmitato, uno degli acidi grassi che vengono metabolizzati rapidamente. Le aree rosse illustrano la distribuzione del carbonio 11 inalato come monossido di carbonio. In un cuore normale *(a)* i ventricoli sono di un verde brillante. Le aree verdi sono abolite nell'immagine TEP di un paziente con infarto miocardico laterale *(b)* e in un paziente con infarto miocardico anteriore *(c)*. La struttura verde che appare a sinistra del cuore in *b* e in *c* è il fegato.

tività in funzione del tempo, si può ottenere un indice della frazione netta di estrazione del palmitato comprendente ^{11}C in lipidi neutri. In studi compiuti su cuori di animali perfusi è possibile controllare indipendentemente le richieste di flusso e le richieste metaboliche del cuore. In condizioni in cui il flusso è mantenuto costante, ma in cui le richieste metaboliche sono modificate, il tasso di declino della radioattività nel cuore verso la fine della curva tempo-radioattività riflette quantitativamente variazioni nel metabolismo del muscolo cardiaco. Tali osservazioni suggerirono che la diminuita assunzione di acidi grassi liberi indotta dall'ischemia dovrebbe essere facilmente rivelabile dalla TEP in animali da esperimento e in pazienti umani.

Quando ad animali da esperimento si somministra per endovena palmitato con ^{11}C, il muscolo cardiaco accumula in modo omogeneo il materiale marcato. Quando però il muscolo cardiaco è reso ischemico, le immagini TEP presentano una marcata diminuzione nell'assunzione di palmitato con ^{11}C. La distribuzione del palmitato e la misura dei difetti nella sua assunzione osservabili nelle immagini corrispondono da vicino a criteri istologici ed enzimatici di lesioni irreversibili del muscolo cardiaco osservate in esami *post mortem.*

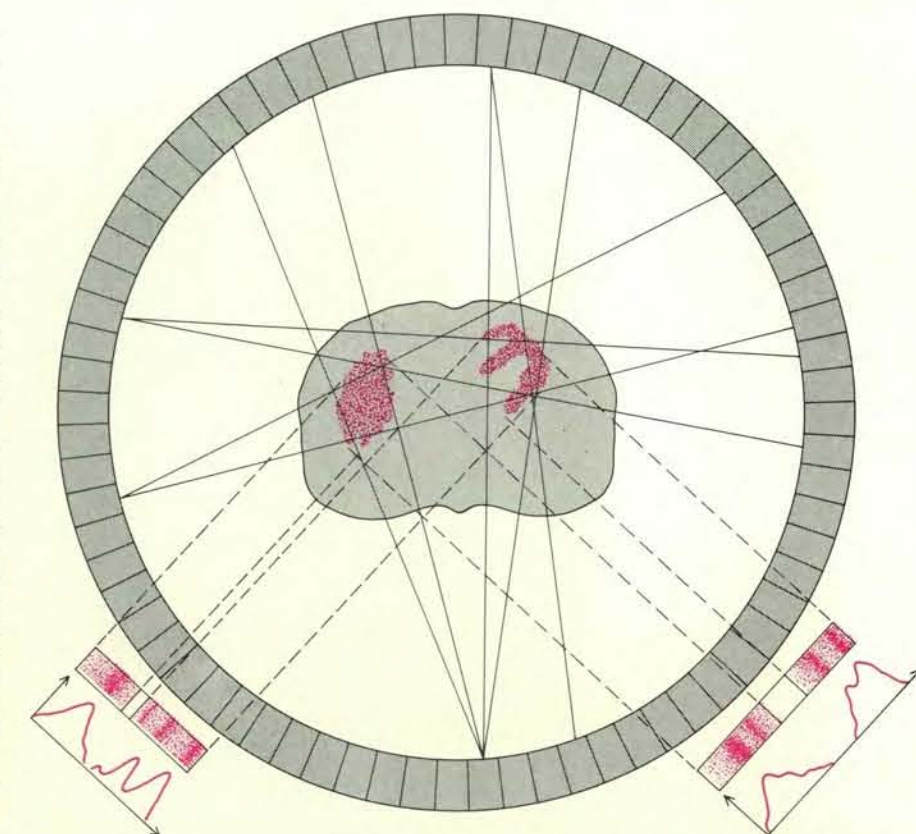
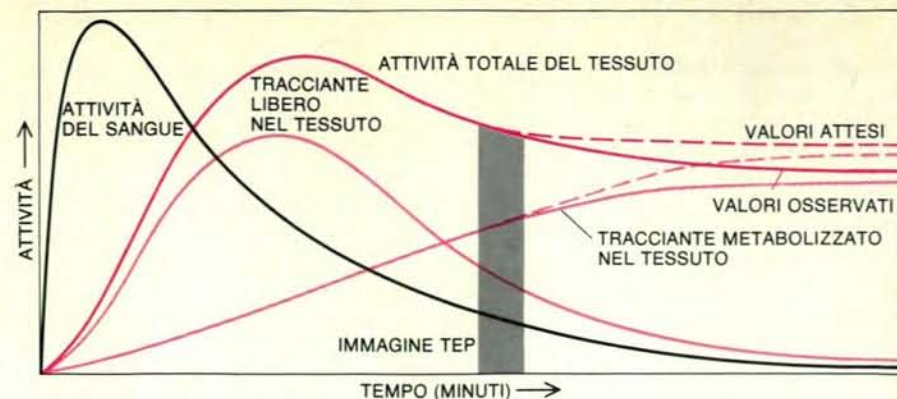
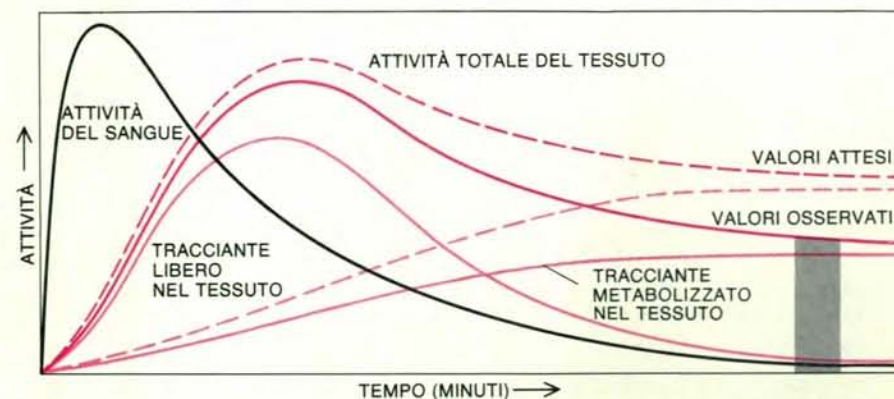
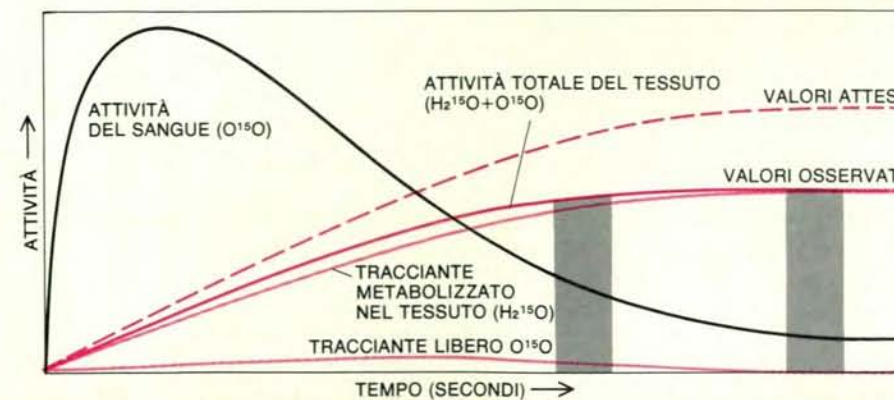
Oltre a evidenziare lesioni del cuore irreversibili, gli studi TEP con palmitato marcato con ^{11}C possono rivelare depressioni del muscolo cardiaco quando il flusso nelle arterie coronariche è stato sottoposto a un'occlusione transitoria. Studi tomografici sequenziali permettono di distinguere fra lesioni reversibili e irreversibili; il muscolo cardiaco leso in modo reversibile non riesce inizialmente ad accumulare palmitato marcato con ^{11}C, ma successivamente, una volta che il flusso sanguigno è stato ristabilito, torna ad accumulare il materiale in modo normale.

Fra soggetti con affezione coronarica conclamata o sospetta, le immagini TEP dimostrano che la distribuzione di palmitato con ^{11}C somministrato per endovena rimane omogena in tutto il ventricolo del cuore. D'altra parte, in pazienti che hanno sofferto in passato di un infarto del miocardio, un episodio che ha come conseguenza la morte di alcune cellule del muscolo cardiaco, le immagini TEP rivelano regioni di diminuita assunzione di palmitato in localizzazioni che corrispondono alle sedi dell'infarto iniziale (quali sono definite dall'elettrocardiografia). Inoltre la misura dell'infarto subito da questi pazienti e rivelato dalla TEP corrisponde esattamente all'estensione dell'infarto stimato sulla base di misurazioni biochimiche.

La TEP eseguita con palmitato contenente ^{11}C ha aiutato anche a chiarire il meccanismo responsabile di una menomazione globale, o generalizzata, della funzione di pompaggio del cuore in pazienti con infarto acuto del miocardio. Secondo alcuni autori, le anormalità regionali che si osservano nel moto della parete cardiaca potrebbero essere dovute primariamente a un ritardo nella conduzione degli impulsi elettrici che eccitano il muscolo cardiaco. Secondo un'altra ipotesi, le anormalità rifletterebbero alterazioni delle funzioni biochimiche e meccaniche del muscolo cardiaco stesso. Nei pazienti da noi studiati, la menomazione globale della funzione ventricolare e la localizzazione di anormalità nel moto della parete cardiaca sono entrambe strettamente correlate all'alterazione del metabolismo rivelata dalla TEP nel muscolo ventricolare. È chiaro che per lo meno in questi pazienti le anormalità del moto della parete sono attribuibili principalmente all'alterazione del metabolismo regionale nel muscolo cardiaco indotta dalla diminuita perfusione anziché ad anormalità nella conduzione di impulsi elettrici.

L'esperienza con palmitato contenente ^{11}C fornisce un esempio della potenzialità della TEP nell'accertamento del metabolismo regionale nel muscolo cardiaco. Altri agenti, come il desossiglucosio marcato con fluoro 18, sono stati usati per ottenere immagini qualitativamente simili a quelle fornite dal palmitato con ^{11}C. È però un po' difficile interpretare con sicurezza le immagini fornite dal fluoro 18, in quanto i composti del fluoro presentano spesso proprietà metaboliche diverse da quelle esibite da metaboliti presenti in natura.

Poiché i traccianti usati negli studi TEP del metabolismo presentano un comportamento complesso, per estrarre i valori del metabolismo dai dati quantitativi forniti da immagini TEP in congiunzione con la radioattività misurata in campioni di sangue arterioso o venoso estratto da un braccio o da una gamba del soggetto, è spesso necessario ricorrere a modelli matematici e schemi di acquisizione di dati abbastanza elaborati. Tali misurazioni richiedono importanti presupposti. Innanzitutto si suppone che la sostanza marcata sia trasportata e metabolizzata nello stesso modo e con la stessa rapidità della sostanza non marcata di cui si vuole ricostruire il comportamento. Questo presupposto viene chiaramente soddisfatto quando una sostanza come il glucosio viene marcata con carbonio 11. Non viene invece soddisfatto nel caso di analoghi, in cui per esempio il fluoro 18 sostituisca atomi di ossigeno; in questo caso devono essere applicati appropriati fattori di correzione.

Un secondo presupposto è che il tracciante metabolizzato rimanga nell'organo in studio durante il periodo della misurazione. Nel caso di un substrato come il glucosio marcato con carbonio 11 ci si può attendere che la sostanza rimanga nel cervello, per esempio, per la durata di tempo necessaria per la scansione di

La proiezione all'indietro di profili convoluti ricostruisce la distribuzione della radioattività all'interno del soggetto in una sezione scelta del corpo. Qui la sezione prescelta passa per il cuore e per il fegato; la colonna vertebrale del soggetto è in basso. Nella convenzione TEP il soggetto viene visto dai piedi, cosicché la struttura in forma di *M* a destra è il cuore e il fegato si trova a sinistra. La distribuzione è quella tipica che si ottiene dopo l'iniezione nel paziente di palmitato contenente ^{11}C. Il palmitato viene trasportato al cuore e al fegato, dove fornisce energia.

L'interpretazione di misurazioni TEP dipende in modo critico dall'uso di modelli matematici che mettono in relazione la concentrazione di particolari traccianti osservati dalla TEP in siti di tessuti specifici con un particolare processo metabolico. Si richiedono due misurazioni simultanee: il livello di attività dell'indicatore radioattivo nel sangue, determinato attraverso il prelevamento di un campione in un'arteria o in una vena periferica, e l'attività nel tessuto bersaglio, determinata dalla TEP. L'attività nel tessuto comprende quella del tracciante metabolizzato e quella del tracciante non metabolizzato. Poiché la tecnica TEP non può distinguere fra le due componenti, assume importanza critica un modello matematico. Per esempio, quando il metabolismo è misurato con un substrato marcato, come glucosio contenente ^{11}C, l'attività nel tessuto dev'essere misurata molto presto per evitare una deviazione dai valori attesi in conseguenza dell'eliminazione dal tessuto del tracciante metabolizzato. Gran parte dell'attività misurata è dovuta al tracciante non metabolizzato, anche se si è tenuto conto dell'attività del sangue nella regione in studio. Un buon modello matematico consente al ricercatore di determinare il metabolismo locale.

Il metabolismo dell'analogo del substrato ^{18}F-2-desossi-D-glucosio differisce dal glucosio normale sotto due aspetti importanti: non viene assunto altrettanto avidamente dalle cellule dei tessuti e una volta assunto non viene eliminato in alcuna forma. Grazie a quest'ultima proprietà la misurazione TEP può essere ritardata, minimizzando così il contributo radioattivo del tracciante libero nel sangue e nei tessuti. Poiché però l'analogo non è accettato dalle cellule altrettanto facilmente del metabolita normale, l'attività del tessuto osservata si discosta assai dall'attività attesa.

La misurazione del consumo di ossigeno richiede un modello matematico ancora diverso. Poiché i tessuti che hanno un'intensa attività metabolica, come il cervello, non immagazzinano quasi ossigeno, l'ossigeno radioattivo (^{15}O) viene metabolizzato rapidamente in acqua ($H_2{}^{15}O$). Perciò la misurazione TEP, quando viene corretta per tener conto dell'attività del sangue, copre soltanto il tracciante metabolizzato. Poiché l'acqua abbandona rapidamente i tessuti, l'attività osservata cade sempre più al di sotto del livello atteso. Per migliorare l'esattezza della misurazione del consumo di ossigeno si possono riprendere varie immagini TEP a intervalli di soli pochi secondi.

un'immagine. Al suo ingresso nelle cellule cerebrali l'^{11}C-glucosio viene fosforilato, una reazione che dà ^{11}C-glucoso-6-fosfato. Poiché il cervello è povero di enzimi in grado di invertire questa reazione ed è impermeabile al glucoso-6-fosfato, il carbonio 11 viene conservato finché l'^{11}C-glucoso-6-fosfato si trasforma in una serie di prodotti di degradazione, i quali vengono infine eliminati dal cervello attraverso il flusso sanguigno venoso. Di conseguenza il carbonio 11 rimane intrappolato efficacemente nell'organo abbastanza a lungo (di solito cinque minuti) perché il metabolismo possa essere misurato mediante la registrazione di immagini TEP. Quando si usano analoghi come il glucosio marcato con fluoro 18, la richiesta di tempo è molto meno rigida perché il processo metabolico si arresta nella fase della fosforilazione. Il fosfato marcato col fluoro 18 rimane catturato irreversibilmente nelle cellule del sistema nervoso.

Un terzo presupposto, negli studi sul metabolismo condotti con la tecnica TEP, è che la quantità del tracciante non metabolizzato, che rimane quindi come tracciante libero nel sangue e nel liquido extracellulare, sia trascurabile o che di esso si tenga conto in modo esatto al tempo della misurazione. Questa richiesta viene soddisfatta nel modo più facile quando i traccianti sono analoghi radioattivi. Poiché simili traccianti rimangono catturati in modo irreversibile nell'organo in studio (e nelle cellule di altre parti dell'organismo), è possibile rimandare la misurazione TEP fino a un momento in cui il tracciante libero nel sangue e nei tessuti sia calato a livelli insignificanti. Quando i traccianti sono analoghi del glucosio è sufficiente un ritardo compreso fra 30 minuti e un'ora. Quando invece non sono analoghi, ma sono veri substrati metabolici, un tale ritardo non è possibile in quanto i metaboliti marcati cominciano a essere eliminati dal tessuto circa cinque minuti dopo il tempo dell'iniezione. Il progetto dell'esperimento deve quindi tener conto esplicitamente della quantità di tracciante libero presente nel sangue e nei tessuti. Si soddisfa in parte questa richiesta misurando il volume regionale della massa sanguigna in connessione con la misurazione del metabolismo regionale. Conoscendo il volume regionale di sangue è possibile calcolare la quantità di tracciante presente non solo nel sangue, ma anche nel tessuto stesso.

Un'alternativa alla richiesta di tener cônto della quantità di tracciante libero nel sangue e nei tessuti consiste nel seguire l'uscita del tracciante metabolizzato dall'organo in studio ripetendo la misurazione del metabolismo varie volte nei primi cinque minuti dopo l'iniezione del tracciante. Il tasso metabolico calcolato sulla base di successive misurazioni diminuirà ovviamente col tempo, e dal tasso di diminuzione è possibile inferire il vero tasso metabolico. In studi preliminari un tale procedimento sembra del tutto realizzabile. Noi abbiamo verificato que-

sto procedimento sia con il glucosio contenente ^{11}C sia con l'ossigeno contenente ^{15}O. Quest'ultimo test è particolarmente convincente grazie alla velocità con cui l'ossigeno 15 viene metabolizzato in acqua ed eliminato dai tessuti. In entrambi i casi abbiamo ottenuto valori eccellenti per il vero tasso con cui le sostanze marcate vengono utilizzate dall'organismo. Nei primi esperimenti ci siamo limitati a immagini TEP di una singola «fettina» di tessuto. Oggi risultati ancora migliori possono essere ottenuti da sistemi TEP avanzati in grado di analizzare con rapidità varie sezioni simultaneamente.

Con sistemi TEP veloci in grado di analizzare simultaneamente varie sezioni di tessuto è facile anche stabilire la quantità di tracciante presente nel tessuto che rimane non metabolizzato. I dati necessari possono essere tratti da immagini TEP in sequenza dell'attività regionale nei tessuti in funzione del tempo nel corso di una lenta infusione endovenosa di tracciante. Se un tale metodo funziona come ci si può attendere, esso dovrebbe rivelarci qualcosa di più del semplice tasso di utilizzazione locale di un substrato come il glucosio o il palmitato. Dovremmo essere in grado di stabilire anche la concentrazione del substrato nell'organo in studio, la sua distribuzione spaziale e la durata della sua permanenza. In pazienti umani tali determinazioni sarebbero estremamente importanti per la comprensione ultima di processi patologici oltre che per la loro diagnosi e il loro trattamento.

Benché gli studi sul metabolismo con analoghi substrati marcati con isotopi radioattivi anziché con i veri substrati abbiano il vantaggio che gli analoghi vengono fissati in modo irreversibile nei tessuti una volta che siano stati metabolizzati, il metodo fondato sugli analoghi presenta anche svantaggi. Poiché gli analoghi non sono biochimicamente identici ai substrati naturali, le differenze rispetto ai tassi metabolici normali devono essere determinate sperimentalmente. Inoltre, nel metabolismo di analoghi sussistono forti differenze fra varie specie animali. Altre differenze sono probabili se l'organo in studio è malato. Altre difficoltà sono connesse con il tempo e con i livelli delle dosi di radiazioni. Quando si usano analoghi di substrati deve trascorrere un periodo compreso fra 30 minuti e un'ora prima di poter compiere la misurazione TEP, per dar tempo al tracciante non metabolizzato di essere eliminato dal sangue e dai tessuti.

Questa richiesta di tempo rende difficile lo studio di fenomeni che si esauriscono rapidamente, come un attacco epilettico, che può durare solo pochi minuti. Si richiedono anche traccianti con lunghi periodi di dimezzamento; in caso contrario, infatti, si deve somministrare al paziente una dose piuttosto elevata della sostanza radioattiva. Infine, analoghi di substrati, se somministrati in quantità eccessive, possono alterare il normale metabolismo della cellula. Si deve perciò aver cura che l'analogo marcato radioattivamente sia relativamente esente da materiale non marcato.

Oltre a determinare i tassi di processi continui come il metabolismo, la tecnica TEP può misurare quantitativamente la composizione chimica di tessuti dell'organismo per mezzo di una scelta idonea di composti marcati che si distribuiscono rapidamente fra il sangue e l'organo in studio. Per esempio, il contenuto di anidride carbonica del cervello della scimmia reso è stato misurato facendo inalare all'animale anidride carbonica marcata con carbonio 11. Le misurazioni furono eseguite con varie concentrazioni di anidride carbonica nel sangue ed evidenziarono il rapporto atteso fra il livello di anidride carbonica nella massa sanguigna arteriale e il livello nel cervello. Questi esperimenti illustrano anche le potenzialità uniche della TEP per determinare quantitativamente l'acidità del tessuto cerebrale in vivo. Tali misurazioni di livelli di acidità dovrebbero essere di grande valore nella valutazione del danno cerebrale subito in conseguenza di colpi apoplettici o traumi fisici.

PET VI (da *Positron-emission tomography*) è il modello più recente, progettato e costruito alla School of Medicine della Washington University. Nel cerchio vi sono i rivelatori di fotoni che registrano il passaggio dei raggi gamma dovuti all'annichilazione di un positone e di un elettrone.

Il piano generale della misurazione di anidride carbonica nel cervello può essere esteso a esplorare la risposta di pressoché ogni organo a un qualsiasi farmaco che possa essere marcato con un radioisotopo. Un'applicazione ovvia consiste nel misurare i livelli del farmaco in vari organi. Gli studi TEP dovrebbero aiutare a definire, per esempio, l'efficacia terapeutica di anticonvulsivi nel cervello o della digitale nel cuore. Per la prima volta dovrebbe essere possibile rappresentare graficamente in soggetti umani la risposta di specifici recettori e trasmettitori del cervello a farmaci con caratteristiche di legame specifiche. Tali ricerche sono già state eseguite in animali. C'è ragione di sperare che ricerche simili in esseri umani possano permettere di chiarire l'azione di agenti psicotropi su certe condizioni per le quali non esistono modelli animali soddisfacenti, come la schizofrenia e il morbo di Parkinson.

Lo sviluppo di una misurazione del tutto soddisfacente del flusso sanguigno per mezzo della TEP è rimasto indietro, per varie ragioni, rispetto ad altre applicazioni di questa tecnologia. Innanzitutto, la frazione di estrazione di molti traccianti risente delle condizioni metaboliche del tessuto, oltre che del tempo di permanenza in esso, compromettendo così una misurazione esatta della perfusione locale. In secondo luogo, la fissazione incompleta del tracciante inizialmente estratto dal tessuto in studio e il reingresso in circolo del tracciante oscura il rapporto fra accumulo e perfusione regionale. In terzo luogo, l'assorbimento del radioisotopo da parte di componenti del tessuto solubili in grassi può differire sensibilmente dalla quantità assorbita in componenti solubili in acqua, contribuendo in questo modo a introdurre una variabilità nel rapporto fra perfusione regionale e accumulo del radioisotopo all'interno di una data regione dell'organo.

Oltre a queste complessità biologiche, la quantificazione della perfusione regionale di organi come il cuore e il cervello con radioisotopi è stata limitata da difficoltà che si incontrano nella determinazione esatta della quantità di indicatore radioattivo in regioni specifiche dell'organo. La TEP offre la promessa di superare tali difficoltà, cosicché la perfusione regionale di sangue del cuore è stata stimata per mezzo dell'iniezione endovenosa di isotopi scelti e per mezzo della rivelazione della loro distribuzione mediante la tomografia. Un tracciante usato a questo scopo è l'ammoniaca marcata con ^{13}N ($^{13}NH_3$), un composto che si ritiene si comporti come il potassio e come altri ioni positivi a carica singola. Un altro indicatore promettente è il rubidio 81, un analogo del potassio dalle caratteristiche molto simili. L'ammoniaca viene metabolizzata attivamente e viene incorporata nella glutammina dal muscolo cardiaco, cosicché la sua frazione di estrazione non è costante, ma dipende bensí da fattori metabolici.

In cuori isolati di coniglio perfusi, la ritenzione di ammoniaca contenente ^{13}N è influenzata marcatamente dallo status metabolico del cuore in condizioni in cui il flusso può essere controllato rigorosamente. Nonostante tali limitazioni è stato suggerito che il ridotto accumulo di ammoniaca con ^{13}N nel muscolo cardiaco fornisca una stima quantitativa della diminuita perfusione regionale. L'accumulo netto di tracciante radioattivo può essere però il risultato di vari fenomeni in concorrenza fra loro: una diminuzione della frazione di estrazione può essere associata a ischemia regionale ed essere causata da un alterato metabolismo del muscolo cardiaco, aumenti nella frazione di estrazione possono riflettere un tempo prolungato di permanenza in regioni di scarso flusso sanguigno; mutamenti di permeabilità in zone ischemiche possono influire sulla frazione di estrazione e una diminuita eliminazione del tracciante può riflettere la lentezza del flusso sanguigno stesso. La complessa interazione di questi fenomeni e di fenomeni connessi può limitare l'utilità dell'ammoniaca contenente ^{13}N nella determinazione quantitativa della perfusione regionale, anche se la TEP è in grado di rivelare la distribuzione dell'indicatore radioattivo.

In particolare, la misurazione del flusso sanguigno nel cervello è stata finora un compito superiore alla capacità della maggior parte dei sistemi TEP attualmente in funzione in quanto essi sono troppo lenti per raccogliere nel breve tempo disponibile, che viene misurato in secondi, una quantità di dati sufficiente a una ricostruzione adeguata di una sezione di tessuto. Di conseguenza sono stati studiati metodi alternativi. Un metodo consiste nel costruire un'immagine in equilibrio di acqua marcata con ossigeno 15 ottenuta mentre il soggetto sta inalando anidride carbonica contenente ^{15}O, e nel dedurne il flusso sanguigno. Benché la quantità di acqua marcata con ossigeno radioattivo presente nel cervello sia chiaramente proporzionale al flusso sanguigno, la relazione non è lineare. Inoltre, per ragioni che non comprendiamo ancora, il metodo sembra dare risultati irregolari.

Forse il metodo oggi più promettente è una modificazione di un sistema progettato in origine per la misurazione del flusso sanguigno nei topi e in altri animali da laboratorio di piccole dimensioni. Un indicatore capace di diffondersi liberamente viene infuso per un minuto per via endovenosa, dopo di che l'animale viene decapitato. La quantità di tracciante che risulta presente in sezioni del cervello è determinata sulla base di

autoradiografie ed è connessa alla curva tempo-attività nel sangue arterioso dell'animale sino al tempo della sua morte. Sulla base dei due insiemi di dati si può calcolare il flusso sanguigno. Non dovrebbe essere difficile sostituire immagini TEP ad autoradiografie pervenendo così a determinare il flusso sanguigno in soggetti viventi.

La tecnologia della tomografia a emissione di positoni è ben sviluppata e tanto le sue capacità quanto i suoi limiti sono compresi sempre meglio da parte di ricercatori di biologia e, in misura minore, dai clinici. Molti studi TEP sono stati eseguiti a tutt'oggi, con risultati inaccessibili a qualsiasi altra tecnica. Essi comprendono la determinazione del metabolismo regionale nel cervello e nel cuore, studi della permeabilità dei tessuti, la misurazione dell'estensione di infarti lasciati nel cuore da attacchi alle coronarie e promettenti ricerche nella fisiologia delle psicosi. La grande promessa della TEP per la medicina è quella di consentire di acquisire nozioni su condizioni patologiche che sono difficilmente investigabili o addirittura inaccessibili con qualsiasi altro mezzo. La TEP ci offre uno strumento particolarmente efficace per valutare gli effetti di farmaci su tessuti malati o sede di disfunzioni. Un esempio particolarmente efficace ci è offerto dalla dimostrata capacità della TEP di evidenziare alterazioni metaboliche regionali in pazienti schizofrenici, che la farmacoterapia riesce chiaramente a correggere.

Nell'oncologia clinica la TEP offre la possibilità di considerare le neoplasie come entità metaboliche e di misurare gli effetti del trattamento - si tratti di radiazioni, di somministrazione di ormoni o di chemioterapia - dalle variazioni indotte nel tessuto maligno e dalle reazioni biochimiche dei tessuti normali attorno a esso. Nella pratica attuale la conoscenza delle dimensioni di un tumore maligno e della sua risposta alla terapia si fonda quasi per intero sull'esame morfologico: osservazione diretta, compresa la palpazione, di tumori accessibili o la visualizzazione ai raggi X di quelli non accessibili in altro modo.

Le apparecchiature TEP hanno prezzi molto elevati. Questa tecnologia richiede infatti la presenza in un centro medico di un ciclotrone, di strutture chimiche adeguate e di una o più serie di apparecchiature per la produzione di immagini e per il calcolo, oltre a fisici, chimici, matematici, fisiologi e medici capaci e disposti a svolgere un lavoro di gruppo. Gruppi del genere stanno lavorando oggi attivamente in una quarantina di istituti in tutto il mondo. Il merito dello sviluppo della TEP può essere attribuito in grande misura alla Atomic Energy Commission degli Stati Uniti (oggi sostituita dal Dipartimento per l'energia), che sostenne alcuni fra i primi sforzi in questo campo, al National Heart and Lung Institute, che dette molto presto una valutazione positiva delle potenzialità della TEP nella ricerca sulle malattie del cervello e del cuore, e più recentemente al National Institute of Neurological and Communicative Disorders and Stroke, che assunse l'iniziativa finanziando ricerche TEP in vari centri medici. Riconoscendo l'esistenza di un potenziale mercato, un certo numero di società stanno oggi producendo i ciclotroni e i sistemi completi di produzione delle immagini e di calcolo necessari per la tecnologia TEP.

La risoluzione conseguita dalle apparecchiature più perfezionate oggi in commercio è di circa un centimetro, con esposizioni di un minuto. Pare ragionevole sperare che la prossima generazione di apparecchiature migliorerà la risoluzione di un fattore due riducendo l'esposizione necessaria a soli pochi secondi. L'inclusione dell'informazione sul tempo di percorso dei raggi gamma nel processo di ricostruzione, che richiede l'accurata misurazione del tempo differenziale fra l'arrivo di due fotoni prodotti da un'annichilazione ai rivelatori di coincidenza, promette di migliorare considerevolmente la qualità delle immagini TEP, aumentando il rapporto complessivo fra segnale e rumore. La TEP ha raggiunto una fase in cui il suo futuro in biologia e in medicina sembra ormai assicurato.

TRACCIANTE	VITA MEDIA	SOSTANZA MARCATA	USO
OSSIGENO 15	~ 2 MINUTI	OSSIGENO	METABOLISMO
		ACQUA	FLUSSO SANGUIGNO
		MONOSSIDO DI CARBONIO	VOLUME SANGUIGNO
		ANIDRIDE CARBONICA	FLUSSO SANGUIGNO
AZOTO 13	~ 10 MINUTI	AMMONIACA	PERFUSIONE DI ORGANI
			METABOLISMO
		VARI AMMINOACIDI	METABOLISMO
			PERMEABILITÀ (CERVELLO)
		PROTOSSIDO DI AZOTO	FLUSSO SANGUIGNO
		1,3 BIS (2-CLOROETIL) - -1-NITROSOUREA (BCNU)	DOSAGGIO DI FARMACI IN UN TUMORE
CARBONIO 11	~ 20 MINUTI	MONOSSIDO DI CARBONIO	VOLUME SANGUIGNO
		ANIDRIDE CARBONICA	pH DEL TESSUTO
		VARI ALCOOL	PERMEABILITÀ (CERVELLO)
			FLUSSO SANGUIGNO
		VARI ETERI	FLUSSO SANGUIGNO
			CONTENUTO DI LIPIDI NEI TESSUTI
		ACETATO	METABOLISMO (CUORE)
		PALMITATO	METABOLISMO (CUORE)
		METILALBUMINA	EMATOCRITO DEI TESSUTI
		OTTILAMMINA	MAPPA DEI RECETTORI DI MONOAMMINOSSIDASI (POLMONI)
		GLUCOSIO	METABOLISMO
		2-DESOSSI-D-GLUCOSIO	METABOLISMO
		FENITOINA	DOSAGGIO DI FARMACI NEI TESSUTI (EPILESSIA)
		TIMIDINA	METABOLISMO (TUMORI)
		DOPAMMINA, NORADRENALINA	FARMACOLOGIA DEI NEURO-TRASMETTITORI (CERVELLO)
		FLUNITRAZEPAM, ETORFINA, PIMOZIDE	FARMACOLOGIA DEI NEURO-RECETTORI (CERVELLO)
FLUORO 18	~ 110 MINUTI	2-DESOSSI-D-GLUCOSIO	METABOLISMO
		3-DESOSSI-D-GLUCOSIO	METABOLISMO
		ALOPERIDOLO, SPIROPERIDOLO	FARMACOLOGIA DEI NEURORECETTORI (CERVELLO)

Le sostanze radioattive sintetizzate per studi con la TEP forniscono la possibilità di sondare una grande varietà di processi biologici. Gli isotopi a breve vita dell'ossigeno, dell'azoto e del carbonio sono particolarmente utili per ricostruire il metabolismo dei loro corrispettivi normali, ma possono anche essere incorporati in composti di interesse farmacologico. L'isotopo di lunga vita del fluoro, ^{18}F, viene talora sostituito all'ossigeno in analoghi di metaboliti normali o di farmaci.

Il galleggiamento del nautilo

Questo mollusco riesce a equilibrare il proprio peso e a muoversi con agilità grazie alla sua conchiglia divisa in compartimenti che possono venire svuotati o riempiti d'acqua secondo le necessità

di Peter Ward, Lewis Greenwald e Olive E. Greenwald

Prima della fine del Cretaceo, circa 65 milioni di anni fa, la vita negli oceani era completamente diversa da quella attuale; tra i più importanti abitanti di grandi dimensioni dei mari vi era un particolare gruppo di cefalopodi caratterizzati da concamerazioni presenti nella loro conchiglia; di queste concamerazioni i molluschi abitavano solo la più esterna lasciando libere le altre che essi stessi avevano formato secernendo una serie di divisioni o setti. Ciò che diede loro il predominio sui mari fu lo sviluppo della capacità di realizzare un galleggiamento «neutro»: essi avevano, infatti, una densità, o peso per unità di volume, praticamente uguale a quello dell'acqua marina che li circondava. In particolare i cefalopodi a concamerazioni svilupparono un apparato capace di estrarre l'acqua dai compartimenti interni della conchiglia. Con la conquista di queste facoltà (e anche con l'evoluzione di un getto d'acqua direzionale) i cefalopodi a concamerazioni si emanciparono rispetto ai loro antenati che strisciavano sul fondo e divennero i primi grandi carnivori in grado di nuotare liberamente nel mare.

I cefalopodi a concamerazioni raggiunsero la più grande differenziazione durante il Triassico, il Giurassico e il Cretaceo, da 225 a 65 milioni di anni fa. Verso la fine di questo ultimo periodo il loro numero cominciò a diminuire, forse in relazione con l'incremento di un nuovo gruppo di predatori mobili, i moderni pesci ossei, il cui galleggiamento «neutro» viene mantenuto con un sistema del tutto diverso: essi hanno una vescica natatoria che può venire riempita di gas. Qualunque sia stata la ragione, l'estinzione dei cefalopodi a concamerazioni divenne quasi completa alla fine del Cretaceo. L'unico genere sopravvissuto fino ai nostri giorni è il *Nautilus* il quale ci fornisce l'opportunità di studiare l'antico sistema di galleggiamento. La precisione del sistema di galleggiamento del nautilo è degna di nota: la differenza di peso fra un nautilo adulto del peso di circa 1400 grammi e un volume uguale di acqua marina può essere di un grammo.

Le specie di *Nautilus* tuttora viventi abitano le acque che si estendono all'esterno della barriera corallina che circonda le isole del Pacifico tropicale occidentale. Essi salgono raramente in prossimità della superficie, anzi gli esemplari catturati si trovavano a una profondità di circa 600 metri. Poiché essi vivono al di sotto del normale campo di attività dei sommozzatori, si conosce ben poco a proposito del loro comportamento e della loro ecologia. Ciò nonostante sono state fatte alcune osservazioni. In primo luogo lo stomaco di un nautilo sezionato contiene generalmente pezzi di crostacei che dimorano sui fondali. Inoltre nelle rare occasioni nelle quali il nostro gruppo ha osservato il nautilo in mare, l'animale si trovava sul fondo o vicino a esso. Infine il nautilo è un nuotatore lento; un subacqueo può facilmente tenere dietro a uno di essi. Abbiamo la netta impressione che il nautilo si nutra brucando lentamente lungo i pendii della scarpata continentale.

Le parti viventi del nautilo consistono di due sezioni principali, il capo (più precisamente il cefalopodio) e il corpo. Il capo è coperto dal cappuccio, un tessuto carnoso e resistente che funge da scudo. Il nautilo può ritirarsi dentro la sua conchiglia, lasciando esposto solo il cappuccio.

Le caratteristiche più appariscenti del capo sono i tentacoli, che assommano a più di 90, di gran lunga più di quanti ne abbia ogni altro cefalopode vivente. Ogni tentacolo è alloggiato in una guaina nella quale può essere ritratto e dalla quale può essere tirato fuori. La superficie dei tentacoli manca delle ventose caratteristiche dei tentacoli di altri cefalopodi, ma è rivestita da una sostanza viscosa che aiuta a trattenere la preda.

I tentacoli circondano un massiccio paio di mandibole che assomiglia al grosso becco di un pappagallo. Le mandibole sono abbondantemente calcificate permettendo al nautilo di spezzare persino il più protetto esoscheletro di crostaceo. Contrariamente alla saliva del polpo, un cefalopode affine, la saliva del nautilo non contiene tossine atte a rendere innocua una preda che lotta. Le mandibole del nautilo sono le sue uniche armi di offesa. Al di sotto dei tentacoli e delle mandibole vi è una piega di tessuto chiamata iponomio, usata per la locomozione grazie alla sua possibilità di espellere un getto d'acqua. L'iponomio è in effetti costituito da un paio di lembi muscolari che si arrotolano l'uno contro l'altro formando un imbuto notevolmente flessibile.

Un'altra caratteristica del capo è degna di nota. Come altri cefalopodi il nautilo ha occhi prominenti, uno per lato, ma diversamente dagli occhi degli altri, appartenenti alla stessa classe, gli occhi del nautilo sono scarsamente sviluppati, per esempio non sono dotati di cristallino. Una minuscola apertura lascia entrare la luce, e presumibilmente le immagini si formano come farebbero in una fotocamera a foro stenopeico. La nostra esperienza personale con il nautilo ci suggerisce che i suoi occhi possano servire solamente per percepire dei cambiamenti nell'intensità della luce. D'altro canto i tentacoli sembrano portare cellule sensibili alla presenza di sostanze chimiche; queste cellule sembrano essere un valido sostituto della visione.

L'altra sezione principale di un nautilo vivente, il corpo, comprende un grande sacco che contiene gli organi dell'animale. Una gran parte dello spazio è occupata dagli apparati digestivo e riproduttivo; lo stesso sacco è avvolto dal mantello, un lembo di tessuto che secerne la conchiglia. La parte posteriore di quest'ultimo secerne i setti che dividono la conchiglia in compartimenti. Lo spazio fra il mantello e il sacco è aperto sotto il corpo, dando luogo a un'ampia cavità che comunica con l'iponomio. La cavità contiene quattro ampie branchie (per cui questo gruppo di cefalopodi viene anche detto tetrabranchiati) mentre gli altri cefalopodi hanno solo un paio di branchie e sono quindi chiamati dibranchiati; la cavità contiene inoltre i canali d'uscita degli apparati digerente e riproduttore.

I sessi nel nautilo sono ben distinti: il maschio è leggermente più grande della femmina a causa della presenza di un grande organo, lo spadice, che introduce un pacchetto di sperma nella cavità del mantello della femmina durante l'accoppiamento. Le femmine di molte specie di cefalopodi producono in un anno migliaia di uova; una femmina media di nautilo produce non più di 10 grandi uova.

La conchiglia nella quale il nautilo vive è approssimativamente un cono cavo strettamente avvolto a spirale attorno a se stesso; essa è composta da strati di aragonite, una forma cristallina del carbonato di calcio, alternati a strati di una sostanza proteica che ha una composizione chimica simile a quella presentata dalle nostre unghie.

La divisione interna della conchiglia in compartimenti suggerisce immediatamente che il nautilo possa far variare il suo galleggiamento come un sommergibile. Già nel 1696 Robert Hooke scrisse: «L'animale ha il potere di riempire o vuotare d'acqua ognuno [dei compartimenti], come gli basterà per bilanciare ed equilibrare la posizione del suo vascello, o conchiglia, per la navigazione o il viaggio che intende intraprendere; oppure se esso deve risalire, può svuotare queste cavità dall'acqua o riempirle d'aria.» Noi, comunque, non abbiamo mai trovato un nautilo in acqua poco profonda intento a riempire d'acqua una concamerazione; esso può solo vuotare le concamerazioni. Inoltre esso non le può vuotare che lentamente; il mantenimento del galleggiamento è in realtà uno sforzo che dura tutta la vita. Come l'animale cresce all'interno della conchiglia, esso costruisce dietro di sé un setto, formando quindi una nuova concamerazione. Essa viene inizialmente riempita con un liquido acquoso chiamato liquido camerale, ma che viene poi lentamente levato. Questo fornisce la spinta di galleggiamento di cui l'animale necessita per bilanciare il peso crescente delle parti viventi e della conchiglia.

La rimozione del liquido dalle concamerazioni è effettuata dal sifone, un'appendice di tessuto che attraversa un tubo calcareo che collega la parte posteriore del mantello dell'animale attraverso tutte le concamerazioni della conchiglia, comprese le prime. Al centro del sifone si trova una rete di vasi sanguigni; l'acqua marina circostante esercita una pressione

Nautilo fotografato di notte a una profondità di circa 9 metri vicino alla barriera corallina che circonda l'isola di Nuova Caledonia nel Pacifico meridionale. Il mollusco si libra praticamente senza sforzo perché il suo peso è virtualmente uguale a quello dell'acqua di mare che sposta. A pochi metri da dove è stata scattata la fotografia, la barriera inizia a scendere fino a raggiungere una profondità di alcune centinaia di metri. Il nautilo potrebbe essere migrato verso l'alto durante la notte (periodo in cui è attivo), oppure essersi nascosto vicino alla sommità della barriera durante il giorno. Dal momento che non è facile vedere e osservare nautili vivi, disponiamo di una documentazione estremamente scarsa a proposito del loro comportamento. Il subacqueo visibile sulla destra è Pierre Laboute, che ha preso parte alle ricerche svolte dagli autori.

idrostatica sul corpo del nautilo e questa pressione viene trasmessa al sangue che circola in essa. Dunque la pressione del sangue nel sifone è uguale alla pressione sanguigna generata dal cuore più la pressione dell'acqua marina, che aumenta di un'atmosfera ogni dieci metri di profondità. A 400 metri, una profondità tipica per il nautilo, l'interno del sifone avrebbe una pressione di più di 40 atmosfere. In una concamerazione vuota, perciò, dove la pressione è sempre inferiore a un'atmosfera, il tubo calcareo che circonda il tessuto del sifone ha la funzione di impedire al tessuto di esplodere.

Un aspetto della nostra ricerca è stata l'indagine del meccanismo che consente al sifone di estrarre il liquido che inizialmente riempie tutte le concamerazioni. Il liquido assomiglia al sangue del nautilo e all'acqua di mare, dal momento che le concentrazioni di ioni potassio, sodio e cloro sono virtualmente uguali in tutti e tre. È dunque impossibile asserire con certezza se il liquido sia un filtrato del sangue, una secrezione del sifone o del mantello, oppure acqua di mare modificata dall'animale. Tuttavia se le cellule del sifone servissero a trasportare alcuni degli ioni dal liquido ai vasi sanguigni situati al centro del sifone, la concentrazione di ioni nel liquido sarebbe inferiore a quella nel sangue. Come risultato, l'acqua nel liquido passerebbe per osmosi dalle concamerazioni al sangue; da qui potrebbe poi essere estratta dai reni dell'animale.

Questa ipotesi non spiega però come la pressione idrostatica profonda nell'oceano favorisca la tendenza dell'acqua a passare dal sifone alle concamerazioni. Certamente anche se il sifone può ridurre la concentrazione di ioni all'interno del liquido camerale fino a zero e per mezzo di ciò massimizzare il passaggio per osmosi dell'acqua al di fuori delle concamerazioni, è possibile calcolare che, già a una profondità di poco superiore ai 240 metri, la pressione idrostatica diventa abbastanza forte da spingere l'acqua nella direzione opposta. Dal momento che i nautili si trovano fino a profondità di 600 metri con concamerazioni completamente vuote, eravamo quindi favorevoli a escludere il semplice meccanismo osmotico. Volevamo dimostrare anche sperimentalmente che esso non era sufficiente, ma non potendo svolgere esperimenti in aperto oceano abbiamo praticato un piccolo foro attraverso la conchiglia del nautilo, perforando una concamerazione pressoché vuota. Abbiamo poi estratto tutto il liquido presente e abbiamo aggiunto cinque millilitri di una soluzione la cui concentrazione di ioni era superiore a quella del sangue del nautilo. Se il movimento dell'acqua attraverso il sifone fosse governato solo da forze osmotiche, l'acqua si sarebbe diffusa dal sangue del sifone nel liquido camerale artificiale, ma in realtà abbiamo osservato l'effetto opposto. Il sifone era in grado di estrarre il liquido dalle concamerazioni anche quando la concentrazione di ioni nel liquido era quasi doppia di quella del sangue.

A questo punto abbiamo rivolto la nostra attenzione al lavoro di Jared M. Diamond: negli anni sessanta Diamond e i suoi colleghi della Harvard University hanno dimostrato che la cistifellea del coniglio può trasportare l'acqua contro gradienti sia osmotici che idrostatici. L'acqua passa dalla cavità interna della cistifellea nei vasi sanguigni delle pareti; i ricercatori hanno dimostrato che le cellule delle pareti della cistifellea si sovrappongono sulla superficie interna della cavità, ma che vi sono spazi tra le cellule situate in quelle zone della parete della cistifellea che si trovano più vicine ai vasi sanguigni.

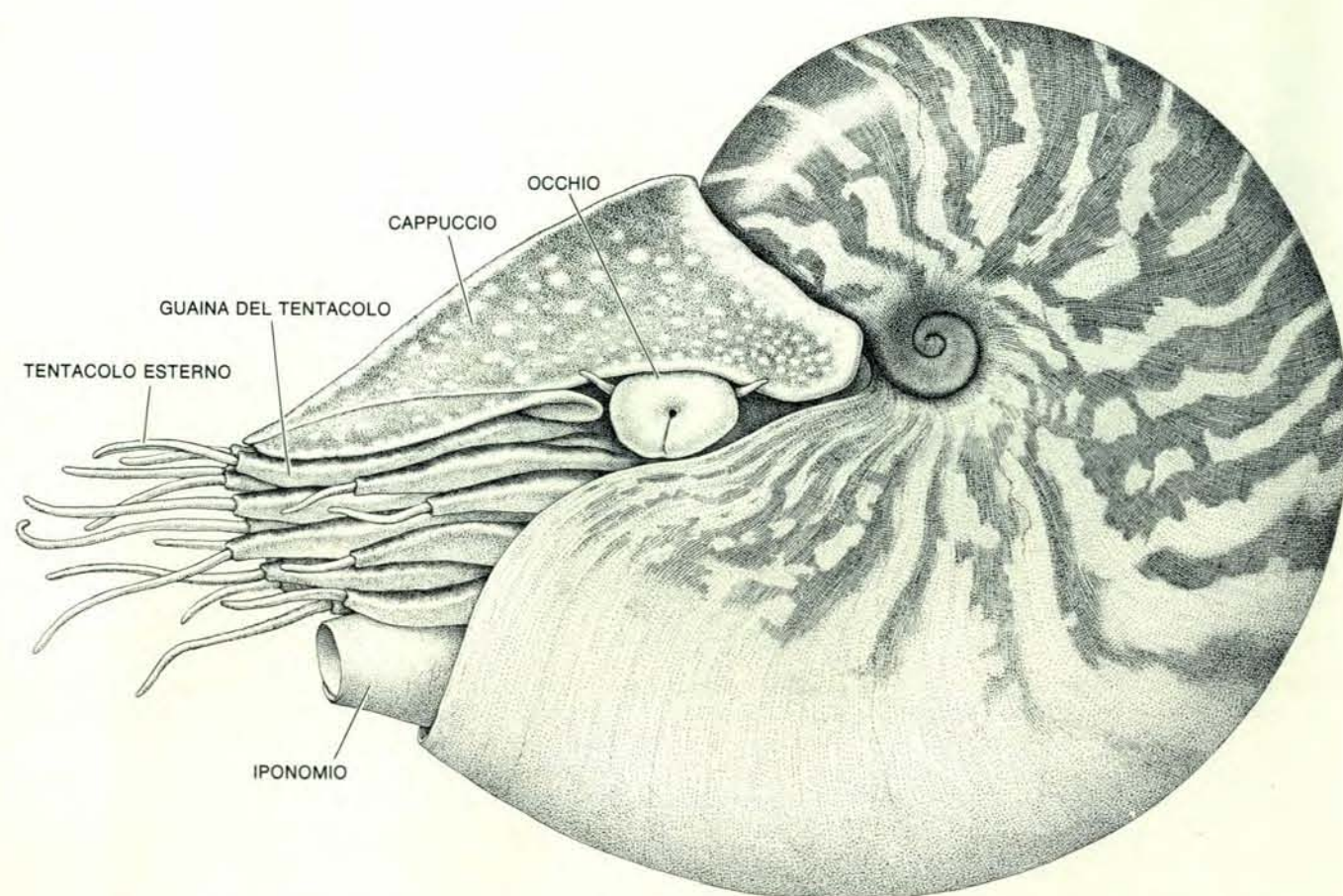
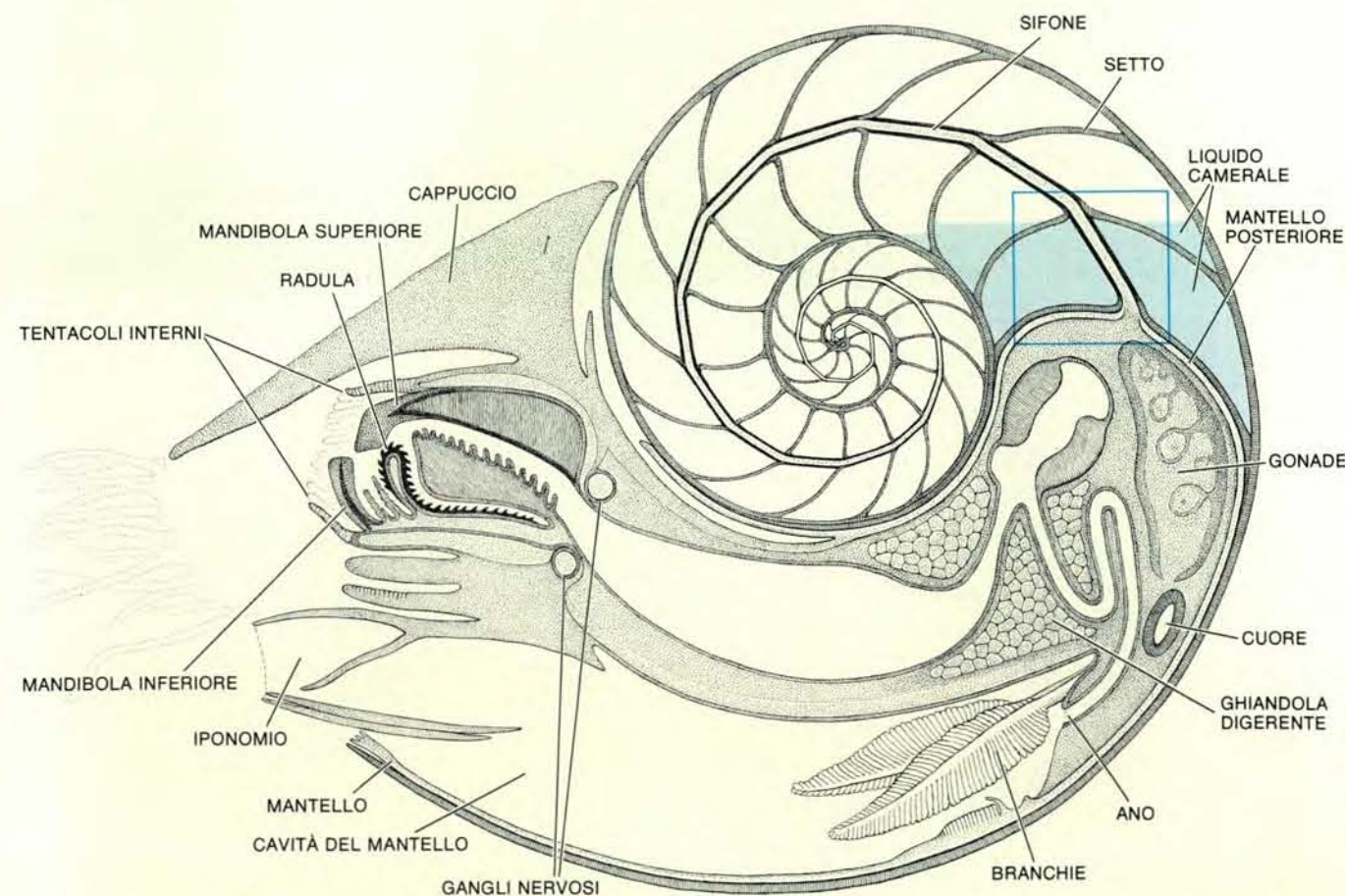
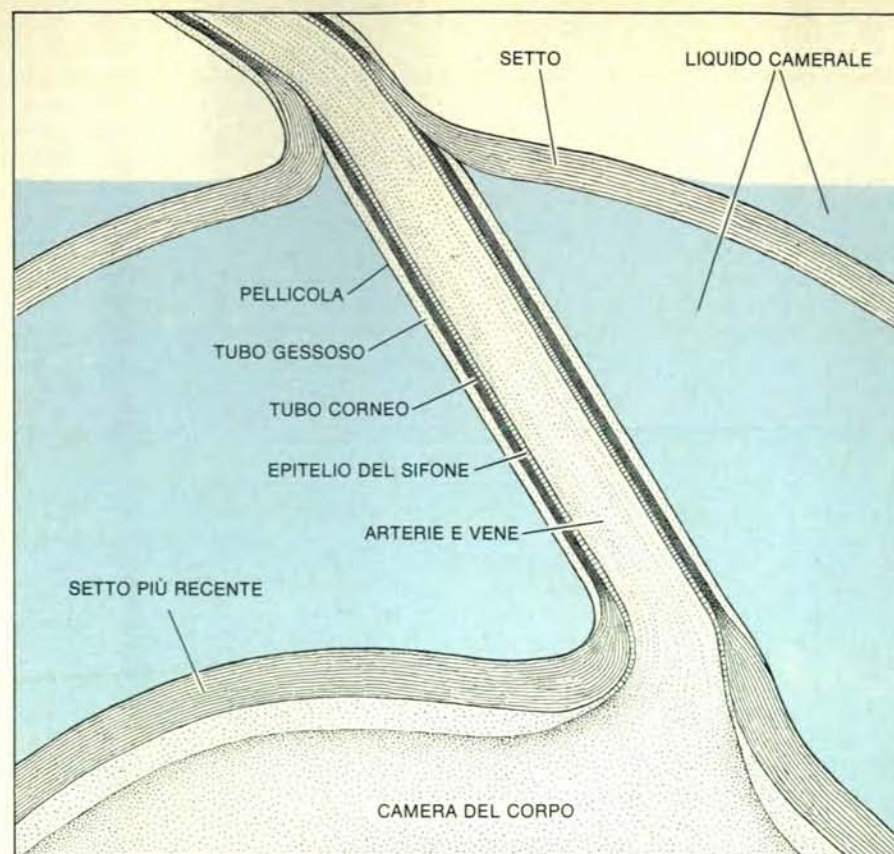
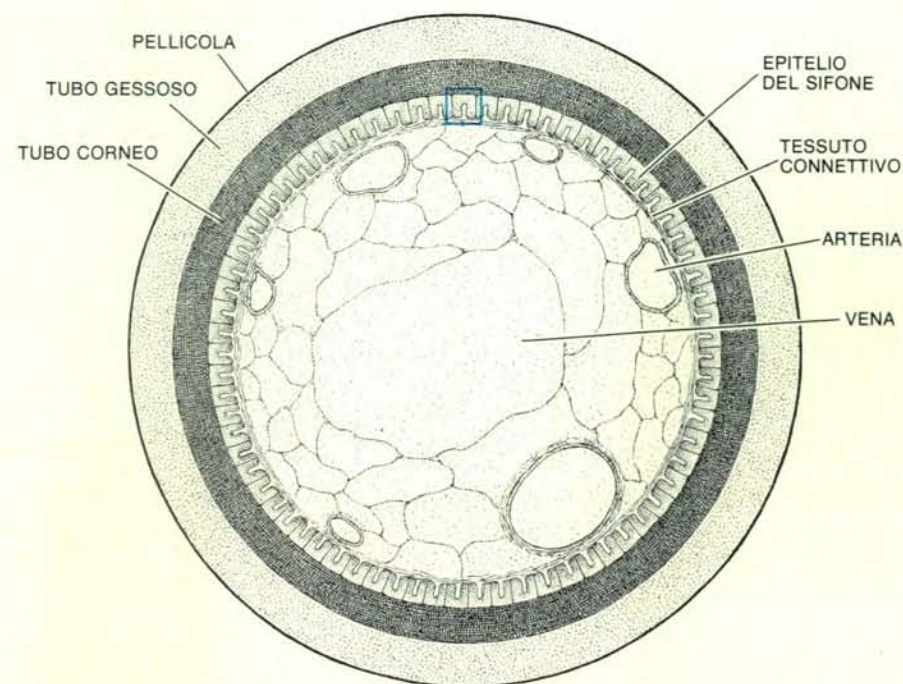
I ricercatori di Harvard hanno suggerito che la membrana cellulare che circonda ogni spazio possa includere enzimi che trasportano ioni attraverso le membrane. Queste pompe molecolari potrebbero dare luogo a elevate concentrazioni locali di ioni negli spazi tra le cellule. Un'analisi matematica ha dimostrato che in queste condizioni l'acqua si diffonderebbe dalla cavità della cistifellea nelle cellule delle pareti e da queste nelle regioni di alta concentrazione di ioni negli spazi tra le cellule. Se le pompe di ioni si trovassero nelle membrane cellulari lungo gli spazi intercellulari chiusi, l'acqua in ingresso scorrerebbe verso gli spazi aperti contro la pressione osmotica o quella idrostatica e, alla fine, entrerebbe nei vasi sanguigni.

Dal momento che questa ipotesi implica la presenza di una pressione osmotica in piccole regioni di uno strato cellulare, è chiamata modello di osmosi locale. Il significato di tale modello rispetto al nautilo è che può spiegare la capacità del sifone di trasportare il liquido camerale sia contro i gradienti osmotici (come quelli da noi riprodotti in laboratorio), sia contro i gradienti idrostatici che l'animale incontra nelle profondità. Il problema è se il sifone ha una geometria cellulare coerente con l'ipotesi dell'osmosi locale.

Evidentemente ce l'ha, poiché nel 1966 Eric J. Denton e John B. Gilpin-Brown della Marine Biological Association di Plymouth, in Inghilterra, hanno pubblicato alcuni disegni del tessuto del sifone che assomigliano in modo sorprendente alle fotografie al microscopio elettronico a basso ingrandimento del tessuto di cistifellea ottenute da Diamond. Denton e Gilpin-Brown hanno in seguito formulato l'ipotesi che l'osmosi locale possa avere luogo nel sifone. Esaminando personalmente il tessuto del sifone anche noi abbiamo notato che i segmenti di sifone ottenuti da concamerazioni che si stavano svuotando o che si erano già svuotate hanno spazi intercellulari sporgenti. Le nostre fotografie eseguite al microscopio elettronico ci suggeriscono comunque che gli ampi spazi intercellulari possono semplicemente servire come canali di drenaggio per liquidi provenienti dalle serie di canali molto più sottili che conducono in spazi più ampi. I nostri motivi per proporre che il liquido camerale venga trasportato prima nella serie di canali più sottili sono basati sull'osservazione che essi sono rivestiti (nel citoplasma che circonda i canali più ampi) da numerosi mitocondri, che riforniscono le cellule di energia. Riteniamo che in questo caso i mitocondri procurino l'energia per il trasporto di ioni o di altre sostanze disciolte.

Tessuti presi da un sifone nella zona in cui attraversa una concamerazione dal setto ancora in costruzione mancano di spazi intercellulari sia ampi che piccoli. D'altro canto la parte del sifone che attraversa concamerazioni che sono state svuotate mantiene la sua rete di spazi. Il sifone dunque sembra essere in grado di svuotare quelle concamerazioni che potrebbero essere riempite dall'acqua spinta verso l'interno dalla forte pressione idrostatica a profondità maggiori di 240 metri.

Nelle concamerazioni vuote, il liquido è stato rimpiazzato da un gas composto essenzialmente di aria da cui è stata estratta la maggior parte dell'ossigeno e a cui è stato aggiunta, probabilmente dalle cellule del sifone, anidride carbonica. Denton e Gilpin-Brown hanno trovato che, quando una concamerazione è stata vuotata, la pressione interna del gas può scendere fino a 0,1 atmosfere. Questo ha fatto loro supporre che il gas non sia trasportato all'interno della concamerazione da un processo che richiede energia; forse lascia il sangue del sifone semplicemente per diffusione.

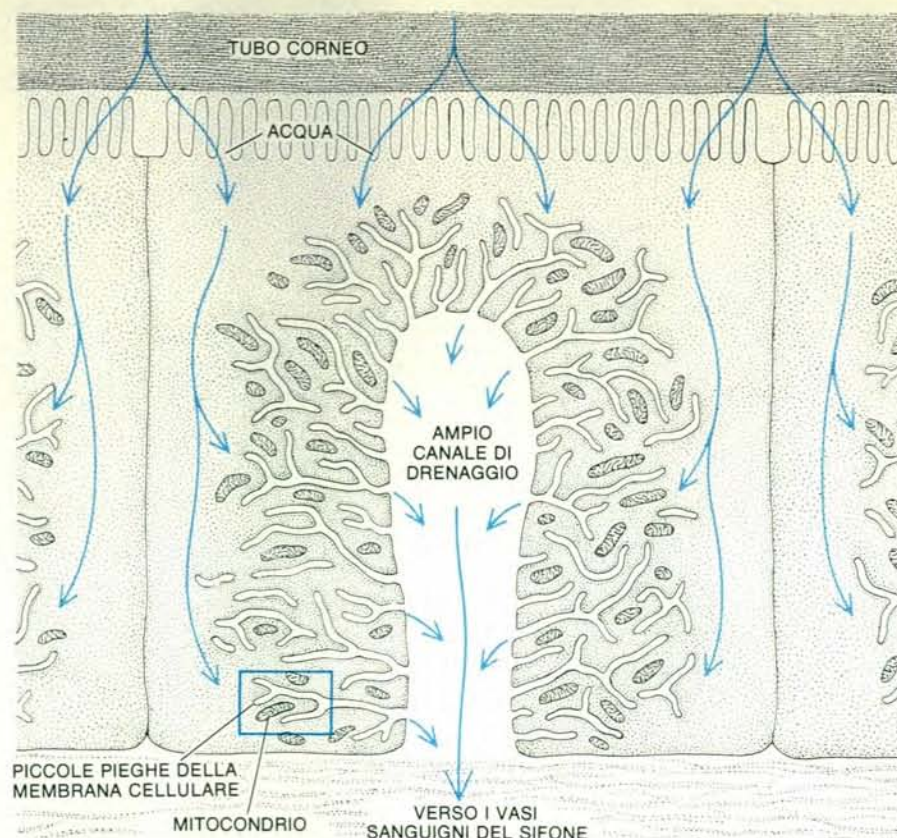
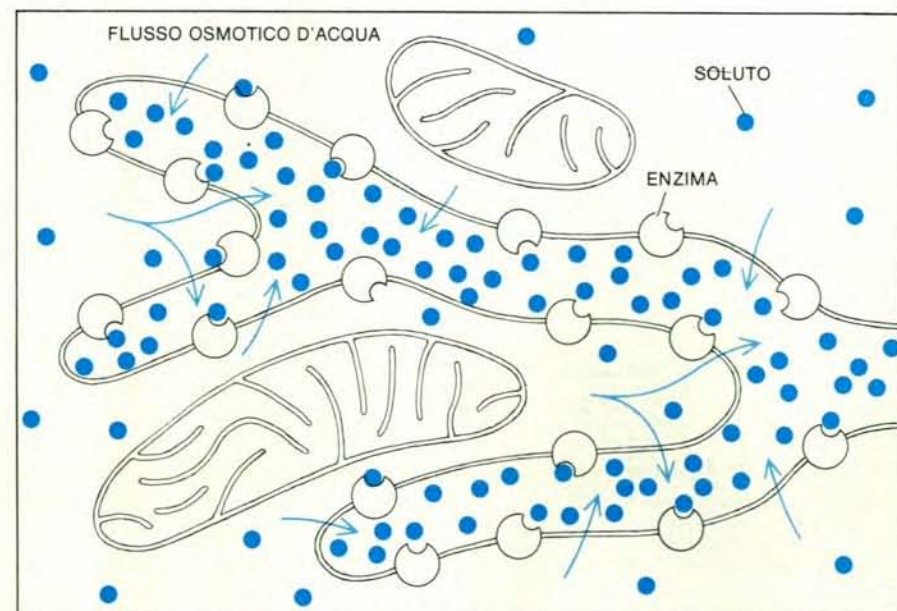
Dal momento che l'acqua marina in cui il nautilo nuota è in equilibrio chimico con i gas atmosferici che si trovano alla superficie dell'oceano e poiché il sangue del nautilo stesso è in equilibrio chimico con l'acqua di mare, non è sorprendente che la pressione dell'azoto e dell'argo nelle concamerazioni piene di gas sia uguale

In queste due figure viene illustrata a grandi linee l'anatomia di *Nautilus macromphalus*, la specie studiata dagli autori: quella di sinistra mostra le caratteristiche visibili esternamente. Il capo dell'animale sporge dalla conchiglia; essa porta più di 90 tentacoli disposti a spirale intorno alla bocca. Gli occhi sono sviluppati, ma la loro visione non è buona come quella di altri cefalopodi; il capo è protetto da un resistente cappuccio carnoso. La figura sulla destra mostra invece la struttura interna della conchiglia e dell'animale: la parte carnosa più esterna è il mantello, un tessuto che secerne la conchiglia. La parte posteriore del mantello secerne la serie di setti che dividono la conchiglia in concamerazioni. Il sifone, un cordone di tessuto rivestito da un tubo calcareo, si estende dalla parte posteriore del mantello attraversando ogni concamerazione. Le due paia di branchie dell'animale sono situate in una cavità posta tra il mantello e i visceri; tale cavità è in comunicazione con l'iponomio, un imbuto flessibile attraverso cui l'acqua viene espulsa in un getto che serve per la locomozione. Quando un nautilo della specie *N. macromphalus* è completamente cresciuto misura vari centimetri e pesa quasi un chilogrammo; a questo stadio la conchiglia ha 30 o più concamerazioni. Il rettangolo in colore mette in evidenza le parti della conchiglia che appaiono nell'illustrazione a pagina 110 in alto.

La più recente concamerazione della conchiglia è quella che si trova immediatamente dietro l'animale e il cui pavimento risulta formato dall'ultimo setto. Inizialmente la concamerazione viene riempita con un liquido acquoso chiamato liquido camerale, ma questo viene in seguito estratto. Man mano che l'animale cresce aumenta la sua capacità di galleggiamento in modo da contrapporsi all'accresciuto peso sia delle parti viventi che della conchiglia. Quando l'ultima concamerazione inizia a vuotarsi, nella precedente si trovano ancora alcuni millilitri di liquido.

Sezione trasversale del sifone, l'organo che gira a spirale attraverso le concamerazioni della conchiglia ed estrae da esse il liquido. Gli strati esterni di questo organo sono porosi e calcarei, ma rivestono un tessuto vivente: uno strato epiteliale nel cui interno scorrono arterie e vene. La pressione del sangue nel sifone corrisponde alla pressione del mare che circonda l'animale, quindi in una concamerazione vuota gli strati esterni del sifone devono impedire ai tessuti interni di esplodere. Il rettangolo colorato mostra la parte del sifone riportata in alto a pagina 112.

alla pressione di quei gas nell'aria a livello del mare. In ogni modo il gas non contribuisce alla spinta di galleggiamento dell'animale, anzi, la diminuisce leggermente poiché dopo tutto il gas ha un peso, mentre il vuoto non pesa nulla. Lo scopo per cui si mette del gas in un pallone è di stabilizzarne le dimensioni e la forma, mentre queste funzioni vengono assolte nel nautilo dalla sua rigida conchiglia.

Un nautilo pesa da un grammo a cinque grammi in acqua di mare, qualunque siano le sue dimensioni, quindi il mollusco conserva il suo galleggiamento «neutro» man mano che cresce. Ciò avviene mantenendo in equilibrio due velocità: quella con cui la conchiglia in accrescimento e le parti molli aumentano di peso e quella con cui il nautilo svuota le concamerazioni della conchiglia dal loro liquido. Per capire come queste velocità siano correlate tra loro, Desmond H. Collins dell'Ontario Museum, G.E.G. Westermann della McMaster University, Arthur W. Martin, Jr., dell'Università di Washington e noi abbiamo svolto alcuni studi, considerando vari aspetti della crescita della conchiglia nel nautilo.

Radiografie di un nautilo in accrescimento dimostrano che la crescita coinvolge una sequenza di processi. Innanzitutto la parte del mantello che riveste la parte anteriore del corpo secerne nuovo materiale per la costruzione dell'estremità aperta della conchiglia. In questo modo questa parte cresce sia in lunghezza che in larghezza, ampliando così la camera dove si trova il corpo e consentendo quindi la crescita delle parti molli dell'animale. Quando l'estremità aperta della conchiglia è divenuta abbastanza estesa, l'animale si sposta in avanti lasciando un certo volume di liquido camerale dietro di sé. Il mantello forma un sigillo tra l'animale e le pareti della camera dove si trova il corpo e secerne un sottile strato di sostanza viscosa per completare il sigillo.

In seguito la parte posteriore del mantello secerne un nuovo setto calcareo, iniziando dai bordi della conchiglia e proseguendo all'interno verso il sifone che si trova al centro. Nello stesso tempo il sifone secerne il proprio rivestimento calcareo. La disposizione delle cellule del sifone nella concamerazione appena formata non è ancora del tipo che riteniamo sia il più appropriato per lo svuotamento di una concamerazione mediante osmosi locale. A questo stadio, comunque, il setto in via di formazione non si è ancora completamente ispessito e il liquido nella concamerazione agisce come una barriera contro la pressione dell'acqua di mare circostante, che si trasmette al setto attraverso il corpo del nautilo. Se il liquido venisse estratto prima che il setto fosse spesso a sufficienza, la nuova concamerazione si spaccherebbe e il sifone esploderebbe a causa della pressione del sangue se il liquido camerale fosse assente mentre il sifone è privo della sua guaina calcarea.

Quando il setto è completato la struttura del sifone è divenuta quella che noi

Lo svuotamento di una concamerazione richiede che le molecole d'acqua attraversino gli strati esterni porosi del sifone e le cellule dell'epitelio che lo riveste. La membrana di ciascuna di queste cellule epiteliali forma una rete di piccole pieghe, che a loro volta comunicano con uno spazio intercellulare più ampio che può essere utilizzato come canale di drenaggio, lasciando entrare l'acqua nei vasi sanguigni che si trovano nel centro del sifone. L'acqua può successivamente venire estratta dal sangue dai reni dell'animale, che la espellono poi nella cavità del mantello. Il rettangolo colorato mostra la parte di cellula epiteliale che viene riportata nell'illustrazione in basso.

Questo disegno schematico della rete di piccole pieghe nella membrana di una cellula dell'epitelio del sifone mostra i dettagli del meccanismo per mezzo del quale gli autori suggeriscono che la concamerazione venga vuotata. Lo svuotamento ha inizio quando le sostanze sciolte nel liquido camerale *(punti colorati)* si diffondono nel sifone; gli enzimi che si trovano sulla membrana delle cellule del sifone pompano allora il soluto dentro alle pieghe. L'energia per il pompaggio viene fornita dai mitocondri, organelli intracellulari che soddisfano le necessità di energia delle cellule. Quando la concentrazione di soluto nella rete di pieghe è abbastanza elevata, l'acqua entra *(frecce)* per osmosi nonostante l'elevata pressione oceanica tenderebbe a far uscire l'acqua dal sifone.

riteniamo appropriata per l'osmosi locale e la concamerazione inizia a vuotarsi. Quando la concamerazione è piena, contiene circa 30 millilitri di liquido. Nel frattempo la concamerazione precedente, e a volte la prima o anche la seconda immediatamente successive, contengono pochi millilitri di liquido; le concamerazioni più vecchie sono del tutto vuote. Nelle concamerazioni in cui il livello del liquido è sceso al di sotto del sifone il liquido camerale può essere estratto lo stesso: esso può raggiungere il sifone per azione capillare lungo la superficie interna della conchiglia.

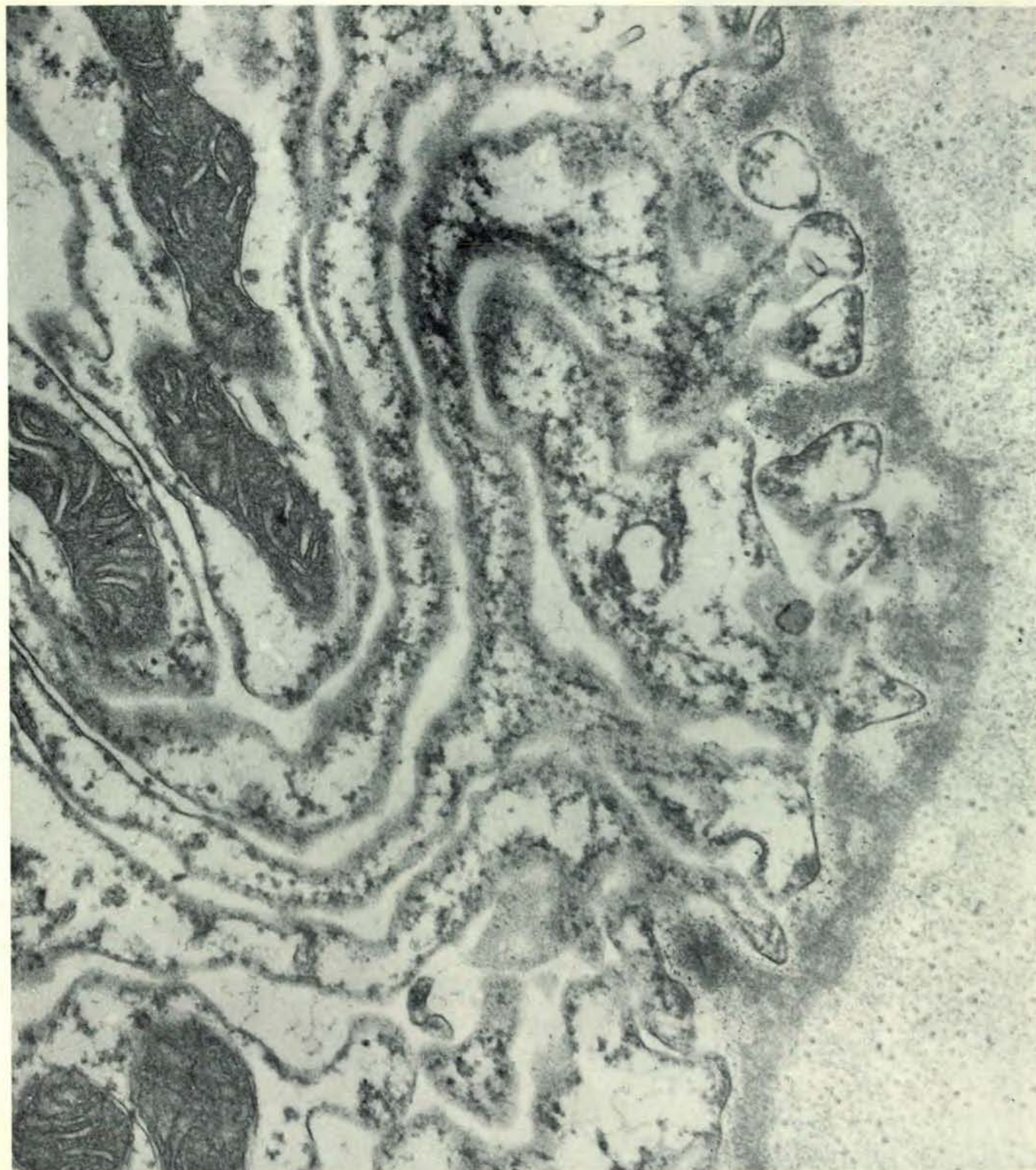
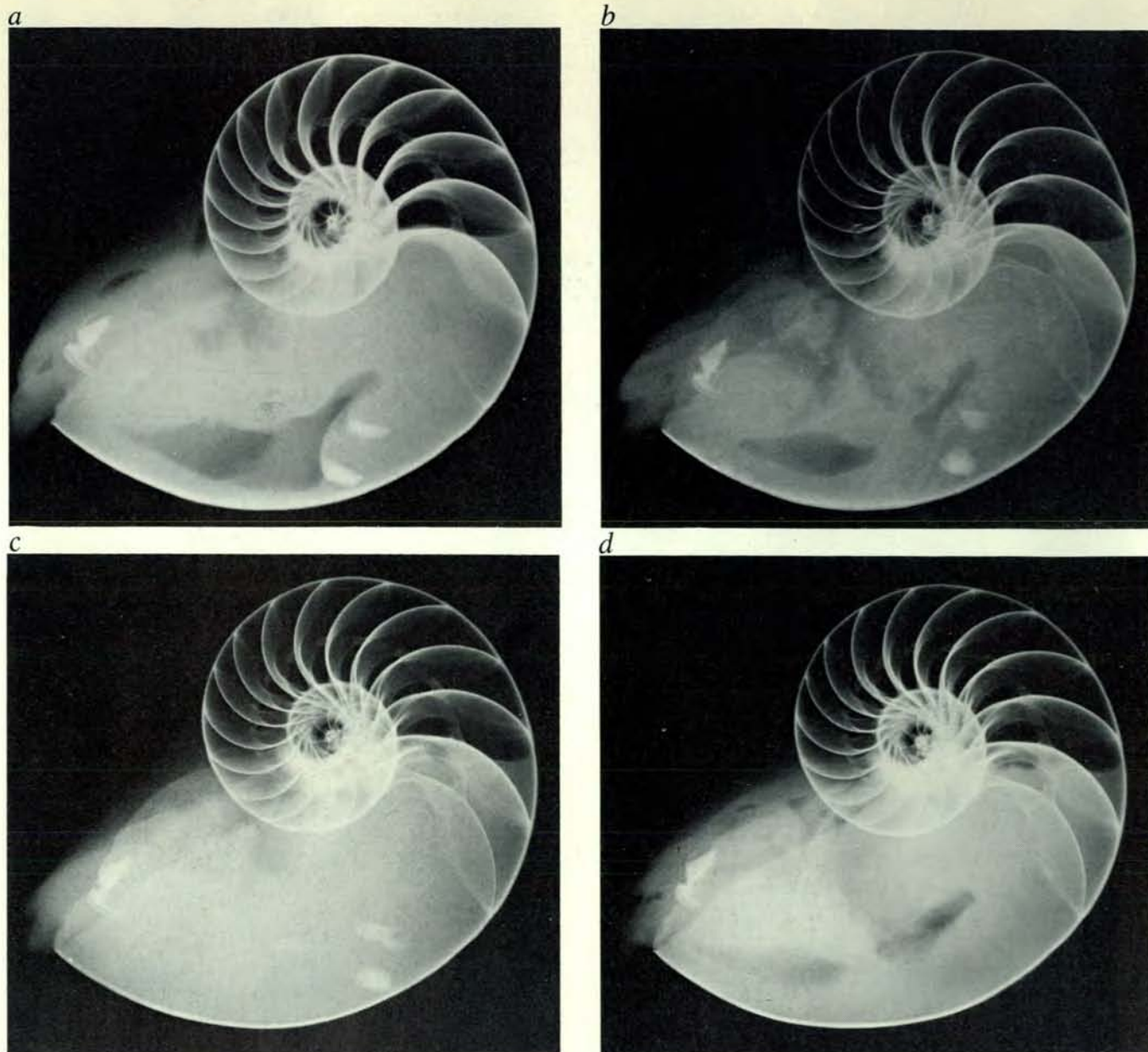
L'estrazione del liquido dalla concamerazione più recente e da quelle adiacenti agisce migliorando il galleggiamento del nautilo che senza questo meccanismo risentirebbe dell'aumento di peso della conchiglia e dei tessuti dell'animale. Questa relazione può essere evidenziata eseguendo il grafico che mette in relazione la lunghezza della camera del corpo con il volume di liquido camerale che si trova subito dietro di essa. Subito dopo che si è formata una concamerazione, la camera in cui è alloggiato il corpo è relativamente corta e il volume del liquido camerale è al massimo, mentre via via che la camera esterna si accresce, il volume di liquido camerale diminuisce. Il ciclo ricomincia quando si forma una nuova concamerazione.

Il nautilo, comunque, non cresce indefinitamente: come molti altri animali esso raggiunge una dimensione definitiva che coincide con il raggiungimento della maturità sessuale. Nel nautilo questo stadio è indicato dalla riduzione degli spazi tra gli ultimi due o tre setti. La riduzione varia notevolmente tra un individuo e l'altro e potrebbe riflettere la capacità del nautilo di effettuare una regolazione finale del suo galleggiamento. In ogni caso la maggiore vicinanza dei setti può essere osservata non solo nel nautilo ma anche nei resti fossili di molti altri cefalopodi a concamerazioni. Ne consegue che molti o forse tutti i cefalopodi a concamerazioni avevano dei limiti definiti nelle dimensioni e non erano a crescita continua.

Quando l'ultimo setto è completato, il nautilo entra nel processo finale che contribuisce a equilibrare la sua spinta di galleggiamento. L'estremità aperta della conchiglia viene allargata e il corpo dell'animale riempie lo spazio; nel frattempo l'ultimo liquido camerale viene estratto dalla concamerazione più recente. La crescita è ora terminata e l'animale con la sua conchiglia misura parecchi centimetri di diametro; la conchiglia ha 30 o più compartimenti e l'animale ha probabilmente almeno tre anni di età. L'allargamento della conchiglia e la secrezione dei setti all'interno di essa ha raggiunto il suo scopo: il nautilo adulto galleggia spontaneamente. Questa situazione si mantiene inalterata se il sifone continua a proteggere la conchiglia contro la pressione idrostatica del mare. Resta ignota la durata della vita del nautilo.

Non abbiamo ancora scoperto se nel nautilo esiste un meccanismo biologico a

retroazione che adegua la crescita della conchiglia con lo svuotamento delle concamerazioni. Se questo meccanismo esiste dovremmo trovare, per esempio, che l'animale non può iniziare la costruzione di una nuova concamerazione se impediamo a quella precedente di vuotarsi. Abbiamo già notato che quando il livello del liquido camerale nell'ultima concamerazione scende sotto il livello del sifone, inizia la secrezione di un nuovo setto.

Inoltre abbiamo dimostrato che il nautilo è in grado di compensare variazioni artificiali nel suo galleggiamento. Nei nostri primi esperimenti ci eravamo trovati di fronte al problema della velocità di svuotamento che era così lenta da poter essere misurata a fatica. Abbiamo allora dato ai nostri animali un «incentivo» a svuotare le concamerazioni sperimentali (quelle in cui effettuavamo misurazioni) aggiungendo liquido in una seconda concamerazione vuota, aumentando dunque il peso dell'animale nell'acqua di mare. In

Questa fotografia eseguita al microscopio elettronico conferma che i mitocondri si concentrano vicino alle piccole pieghe. I mitocondri sono i corpi più scuri; cinque di essi sono visibili, interamente o parzialmente, nella metà sinistra dell'immagine. Il più ampio canale di drenaggio con cui le pieghe comunicano scorre dall'alto verso il basso sull'estrema destra. L'ingrandimento della microfotografia è di 40 000 diametri.

La secrezione di un setto e lo svuotamento delle concamerazioni nella conchiglia possono essere seguiti in questa sequenza di quattro immagini ai raggi X di un nautilo vivente. In *a* il livello di liquido camerale nella concamerazione più recente è sceso al di sotto della base del sifone; nella concamerazione subito successiva resta una quantità di liquido molto inferiore, tutte le altre concamerazioni sono vuote. In *b* (ripresa tre settimane più tardi di *a*) un nuovo setto inizia ad apparire e la concamerazione che delimita è piena di liquido camerale; nel frattempo il livello del liquido nella concamerazione precedente ha continuato a diminuire. Nonostante il fatto che il liquido in quella concamerazione non sia più direttamente a contatto con il sifone, può raggiungerlo attraverso la superficie interna della conchiglia. In *c* (ripresa due settimane dopo *b*) il nuovo setto è quasi completo. In *d* (ripresa due settimane dopo *c*) il nuovo setto è completo. Le due aree più scure in alto a sinistra nella nuova concamerazione sono bolle di gas; la loro presenza indica che essa è in fase di svuotamento. Il nuovo setto è ormai diventato abbastanza resistente da sopportare la pressione dell'oceano anche senza la spinta idrostatica fornita dal liquido camerale. In tutte le immagini il nautilo occupa la concamerazione più esterna della conchiglia. Le due macchie biancastre sulla destra del corpo in *a* sono calcoli renali. Le righe bianche sulla sinistra di ogni immagine sono le mandibole dell'animale; le strutture scure sono parti della cavità del mantello.

tutti gli animali resi più pesanti dal riempimento di una concamerazione abbiamo osservato uno svuotamento molto più rapido della concamerazione sperimentale. Quando gli stessi esemplari sono stati fatti galleggiare in modo anormale svuotando la concamerazione in precedenza riempita artificialmente, abbiamo osservato un abbassamento della velocità di svuotamento. Il più rapido svuotamento osservato in laboratorio è stato di meno di un millilitro al giorno, un valore coerente con la stima secondo la quale il mollusco impiega almeno un mese per vuotare una concamerazione.

Dato che in tutte le nostre osservazioni di nautili in acqua poco profonda non abbiamo mai visto un esemplare aggiungere liquido in una concamerazione precedentemente vuotata, concludiamo che questo procedimento è collegato solamente con la scelta di aumentare o diminuire la velocità di svuotamento delle concamerazioni (e forse la velocità di formazione della conchiglia). Perché allora l'animale attraversa con il sifone anche le prime concamerazioni? La sua vita sarebbe più efficiente se avesse sviluppato al suo posto un meccanismo per rendere ogni concamerazione assolutamente impermeabile dopo che è stata vuotata. Forse la parte del sifone che si estende fino alle prime concamerazioni è un residuo di un organo che poteva riempire e svuotare dal liquido le concamerazioni dei cefalopodi a concamerazioni ancestrali. Quelle creature, ora estinte, sarebbero allora state molto più mobili di quelle in cui si sono successivamente evolute.

Coltivazione di petrolio

Con le usuali tecniche di estrazione, un giacimento petrolifero viene sfruttato in media al 50 per cento. L'aumento dei prezzi del petrolio rende quindi conveniente la coltivazione a giorno o in sotterraneo

di Richard A. Dick e Sheldon P. Wimpfen

Il metodo tradizionale di sfruttamento dei giacimenti petroliferi è notoriamente inefficiente. La quantità di petrolio che, con un pozzo trivellato, si può estrarre da un deposito sotterraneo mediante pompaggio può essere pari addirittura a meno di un decimo del totale e, in media, è notevolmente al di sotto del cinquanta per cento. Ai tempi in cui il petrolio abbondava ed era a buon mercato, nessuno badava molto all'inefficienza di questo tipo di estrazione. Oggi però, in tempi di magra, con le riserve petrolifere che continuano costantemente a ridursi e con i prezzi del petrolio importato che sono arrivati a superare del 1500 per cento quelli esistenti solo una decina d'anni fa, gli Stati Uniti non dovrebbero trascurare più un elemento così estremamente importante. Una buona parte del petrolio rimasto nel terreno dopo il pompaggio può essere estratto con l'adozione di tecniche minerarie.

Forse a qualcuno il concetto di coltivazione mineraria di un giacimento petrolifero potrà apparire radicale. La verità è che negli Stati Uniti l'estrazione del petrolio con tecniche minerarie è vecchia quasi quanto quella attuata per mezzo di trivellazioni. Oltrettutto, all'estero il petrolio è stato e viene tuttora estratto con successo con tecniche di questo genere. Più di un secolo fa il metodo minerario fu applicato all'estrazione del petrolio in due modi diversi nell'Ohio e in California. Nel 1865 a Macksburg (Ohio), quando un pozzo tradizionale cominciò a dare poco petrolio e molta acqua, fu sospesa la perforazione, e al posto del piccolo foro di trivellazione fu scavato un pozzo da miniera di grande diametro. Il petrolio venne poi scremato dalla parte superiore dell'acqua. Purtroppo però il pozzo minerario non produsse molto più petrolio dell'altro.

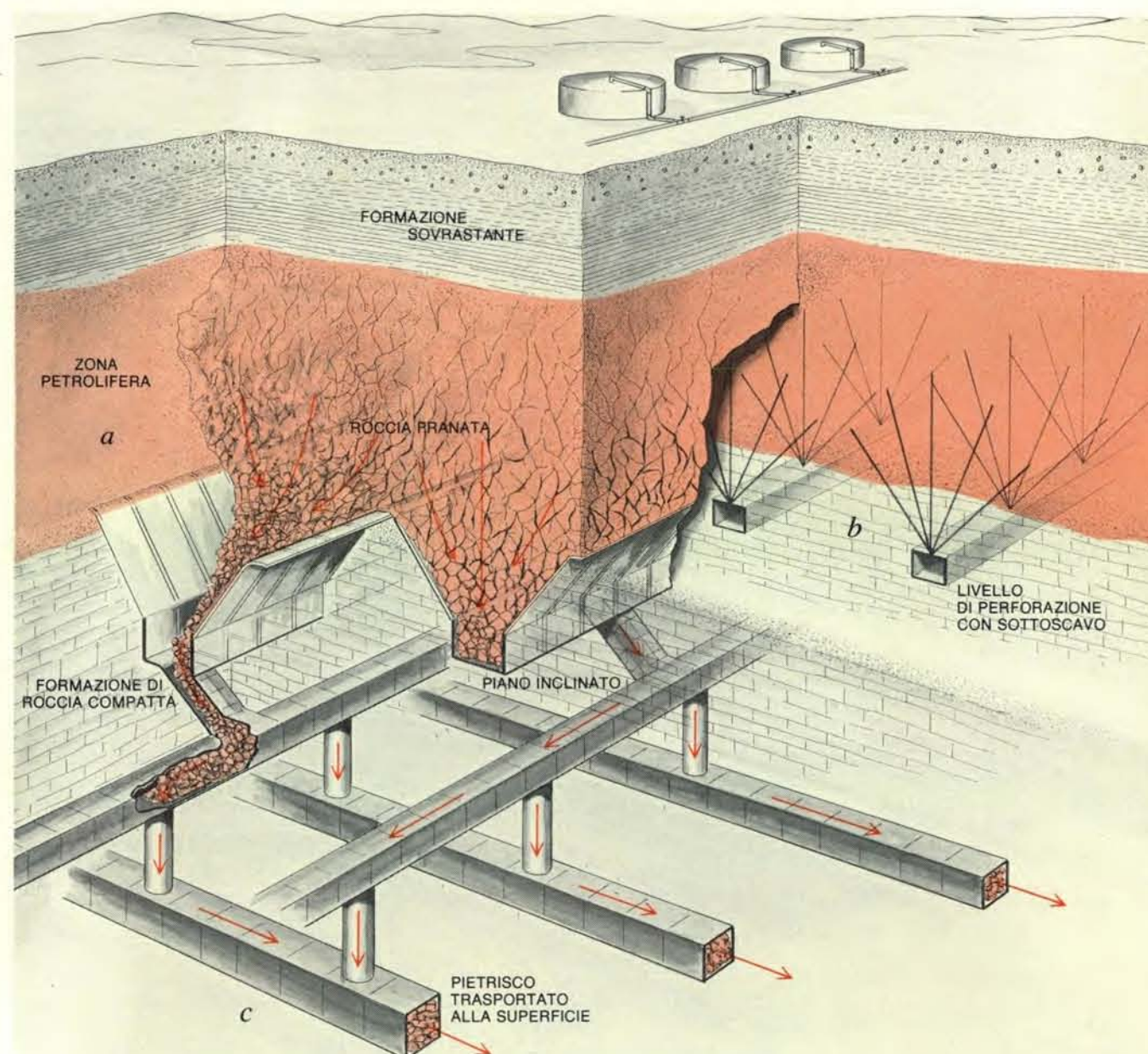
L'anno dopo Josiah Stanford, pioniere dell'industria petrolifera nell'ovest, attirato da infiltrazioni di petrolio sulla Sulphur Mountain, a nord di Ventura (California), scavò nel fianco della montagna una serie di gallerie, tutte inclinate leggermente. Prima che Stanford portasse a termine il suo progetto, erano state scavate 30 gallerie, ognuna delle quali produceva da uno a 20 barili al giorno di petrolio, portato fino alla bocca delle gallerie dalla forza di gravità. Queste e altrettante gallerie scavate ancora più in profondità nella Sulphur Mountain agli inizi degli anni novanta del secolo scorso continuarono a produrre petrolio per parecchi decenni.

In Europa la coltivazione di un giacimento petrolifero con tecniche minerarie è un'arte ancora di più antica data. A Merkwiller-Péchelbronn, nel dipartimento del Basso Reno della Francia orientale, gallerie tipo miniera furono scavate per la prima volta in una formazione petrolifera nel 1735. All'inizio il petrolio filtrava semplicemente in canali scavati sul fondo della galleria, ma nell'Ottocento la raccolta per caduta, o gravità, fu integrata da trivellazioni. Negli anni trenta del nostro secolo la rete di gallerie di questa miniera di petrolio si estendeva ormai su un'area sotterranea di oltre sette chilometri quadrati. Verso la fine del secolo scorso fu intrapresa a Wietze, fra Amburgo e Hannover (Germania Occidentale), un'attività di estrazione del petrolio con metodi minerari che pare si possa mettere sullo stesso piano dello scavo dei pozzi minerari dell'Ohio. Durante la prima guerra mondiale, via via che la domanda di petrolio in Germania aumentava, fu dato maggiore impulso anche agli sforzi di Wietze. Nel 1917 fu scavato un pozzo del diametro di quasi quattro metri fino a una profondità di 150 metri; nel 1918 da quel pozzo furono scremati 127 000 barili di petrolio.

Oggi in tutto il mondo sono il Canada e l'URSS a dedicarsi con maggiore impegno all'estrazione del petrolio con tecniche minerarie. In realtà l'iniziativa canadese riguarda non tanto lo sfruttamento di un giacimento petrolifero quanto quello delle «sabbie bituminose», ma il prodotto finale è il petrolio e le tecniche operative sono istruttive. La sabbia bituminosa (un'arenaria permeata di bitume tenero) si trova fra un basamento calcareo e un terreno di copertura. Il materiale viene estratto con il metodo «a fossa», che comporta l'asportazione di una striscia di materiale di copertura della larghezza di circa 15 metri fino a scoprire lo strato di sabbia bituminosa, che è alto circa 40 metri. La roccia tenera viene raccolta in lunghe andane e trasferita meccanicamente a dei nastri trasportatori che la portano fino all'impianto di trasformazione. Per produrre 45 000 barili al giorno di petrolio greggio sintetico estraendolo dal bitume presente nella roccia, bisogna rifornire l'impianto di trasformazione di 117 000 tonnellate di sabbia bituminosa. (Il greggio viene chiamato sintetico perché viene ottenuto per distillazione del bitume.) L'obiettivo della produzione di 45 000 barili al giorno è già stato raggiunto e superato e un secondo impianto, finanziato da un consorzio internazionale di compagnie petrolifere, non dovrebbe tardare a produrre altri 125 000 barili al giorno.

L'iniziativa sovietica è stata invece una reazione al basso rendimento delle trivellazioni effettuate nel giacimento di Yarega, che si stende a ovest degli Urali vicino a Ukhta, nella repubblica sovietica autonoma dei Komi. Il giacimento, situato circa 198 metri sotto la superficie, è costituito in gran parte di «petrolio pesante», di petrolio cioè troppo viscoso per essere estratto con le tradizionali tecniche di trivellazione e di pompaggio. In effetti è stato calcolato che il pompaggio dall'esterno sfruttava soltanto il 2 per cento del giacimento. Per migliorare la situazione furono scavati pozzi minerari fino a una profondità di oltre 182 metri e a partire da ogni pozzo fu scavata una rete di gallerie orizzontali. Da queste aree di lavoro furono praticati passaggi inclinati diretti verso il basso nella formazione petrolifera e sul fondo di ogni passaggio fu scavata una galleria produttiva del diametro di pressappoco 25 metri. Le trivellazioni produssero allora una serie di aperture orizzontali e leggermente dirette verso l'alto che si irradiavano da ogni galleria produttiva per una distanza pari a quasi 250 metri.

Per favorire il flusso del petrolio pesante verso queste aperture di raccolta, veniva iniettato vapore nella roccia serbatoio attraverso fori verticali trivellati dalle sovrastanti gallerie orizzontali. Il greggio caldo veniva poi riscaldato ulteriormente a un terminal centrale sotterraneo prima di essere pompato alla superficie. Con questa combinazione di trivellazione laterale effettuata dalle gallerie produttive sotterranee e di iniezione di vapore, i sovietici calcolano che sarà possibile recuperare dal 50 al 60 per cento del petrolio del giacimento di Yarega alla media di oltre 6 milioni di barili all'anno. Intanto hanno avviato analoghe operazioni estrattive in un giacimento esaurito situato a nord dei campi petroliferi di Baku, nella repubblica sovietica dell'Azerbaigian, e stanno considerando la possibilità di fare lo stesso a più di 300 giacimenti noti di petrolio pesante in altre località dell'URSS.

Quanto petrolio potrebbe diventare disponibile negli Stati Uniti con il ricorso a tecniche minerarie? È stato calcolato che attualmente il petrolio rimasto nei giacimenti già sfruttati e quello che rimarrà nei giacimenti ora in fase di produzione o tenuti di riserva è pari in totale a più di 300 miliardi di barili. Per avere un'idea migliore dell'entità di questa risorsa potenziale basta raffrontarla con la stima delle esistenti riserve statunitensi di petrolio recuperabile con i metodi tradizionali: meno di 30

Il petrolio presente in una roccia tenera potrebbe essere estratto con la tecnica della coltivazione in sotterraneo. Per prima cosa verrebbe scavata una serie di gallerie comunicanti in una formazione rocciosa compatta che si trova sotto la formazione petrolifera di roccia tenera. La rete superiore di gallerie è collegata a piani inclinati che salgono fino alla base della roccia tenera *(a)*. Perforazioni ed escavazioni per varata e per sottoscavo *(b)* fanno crollare la roccia tenera. I detriti giungono alle gallerie più elevate attraverso piani inclinati, dopo di che passano nella rete inferiore di gallerie *(c)* per essere trasportati, quindi, alla superficie. A questo stadio un impianto di arricchimento separerebbe il greggio dai detriti rocciosi. Questo sistema di coltivazione di un giacimento petrolifero viene detto «per frana a blocchi» *(block caving)*.

miliardi di barili. È ovvio che non tutti i 300 miliardi di barili rimasti nei giacimenti statunitensi possono essere recuperati con tecniche di estrazione di tipo minerario. La cifra di 300 miliardi di barili tuttavia non tiene conto di ulteriori grosse risorse potenziali di idrocarburi.

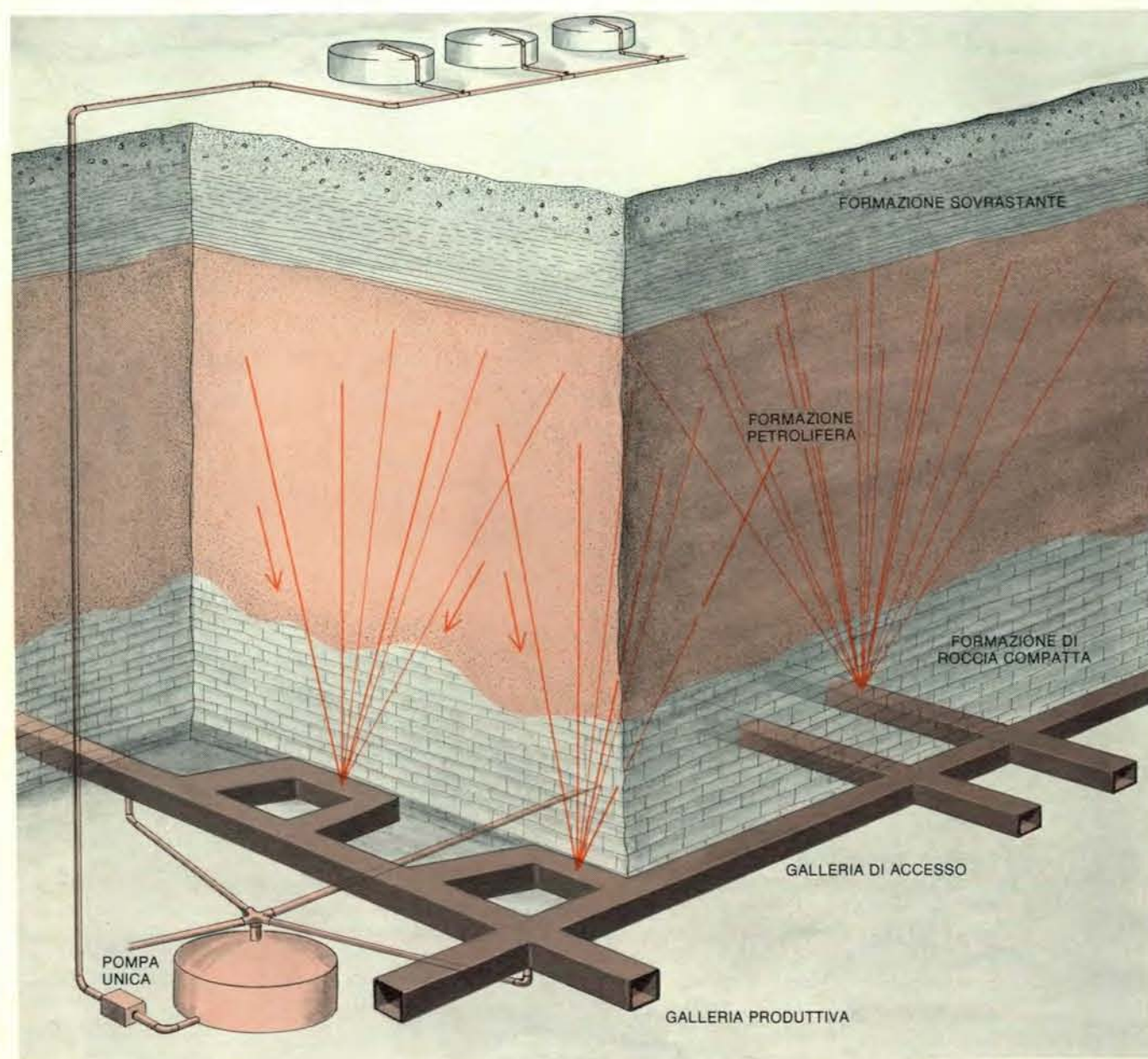
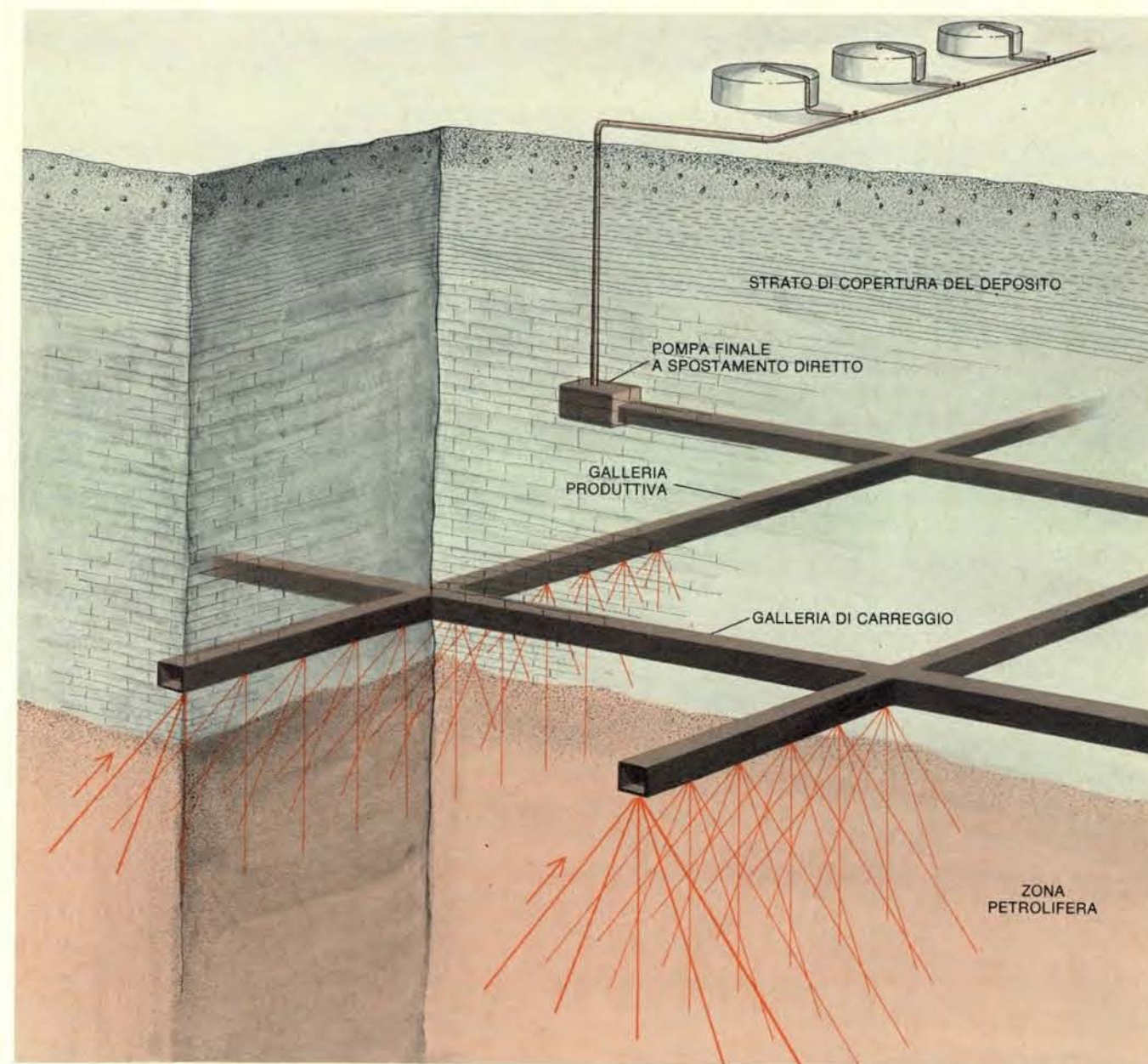
Se alle riserve americane accertate di petrolio pesante simile al greggio di Yarega si aggiungono alcuni depositi diatomitici impregnati di idrocarburi in California e le sabbie bituminose dello Utah (che da sole si ritiene che contengano un potenziale di 30 miliardi di barili di greggio sintetico), si scopre che in potenza sono disponibili altri 200 miliardi di barili di petrolio. Oltre tutto questa cifra esclude il potenziale di produzione di petrolio greggio sintetico dalla formazione di scisti bituminosi del Green River nello Utah, nel Colorado e nel Wyoming. Si calcola che questa formazione costituisca una riserva di 1,8 bilioni di barili, di cui si ritiene che si possano recuperare da 400 a 600 miliardi di barili. Non è tuttavia questa la direzione che seguiremo per lo sfruttamento degli scisti bituminosi con tecniche minerarie.

Non molto tempo fa il Bureau of Mines, un'agenzia del Dipartimento degli interni degli Stati Uniti, ha commissionato due studi di fattibilità, in modo da poter fare una valutazione aggiornata del potenziale di estrazione del petrolio con metodi minerari negli Stati Uniti. L'ultima di tali valutazioni era stata fatta dal medesimo Bureau più di 40 anni prima, quando il petrolio prodotto con metodi tradizionali raramente costava più di 2 dollari al barile e le riserve di petrolio degli Stati Uniti sembravano inesauribili. Per presentare i risultati di questi recenti studi nel loro giusto contesto sarà bene da parte nostra descrivere prima brevemente certi aspetti della tecnologia mineraria e quindi discutere tre criteri generali di fattibilità.

I metodi di estrazione mineraria si possono dividere in tre categorie: coltivazione a giorno, coltivazione *in situ* e coltivazione in sotterraneo. Le operazioni relative alle sabbie bituminose dell'Alberta sono tipiche della coltivazione a giorno; le trivellazioni tradizionali (non solo per il petrolio, ma anche per lo zolfo e per altre risorse di minerali solubili) sono un esempio di estrazione *in situ*. L'estrazione in sotterraneo può assumere molte forme, ma per quel che riguarda il suo potenziale per l'estrazione del petrolio i metodi opportuni sono limitati. Non è affatto pratico, per esempio, scavare una galleria nel sottosuolo e successivamente trasportare alla superficie la roccia petrolifera, così come si fa grosso modo col carbone bituminoso.

Quei serbatoi di petrolio che si trovano in formazioni rocciose sufficientemente compatte da poter dare garanzie di sicurezza per lo scavo di gallerie presentano una bassa proporzione di greggio rispetto alla roccia. D'altro canto, quando la percentuale di petrolio è abbastanza elevata da rendere economicamente allettante lo sfruttamento della roccia serbatoio, la roccia stessa è pericolosamente debole.

Più promettente è il metodo che combina la coltivazione in sotterraneo e quella *in situ*, come nella miniera di petrolio sovietica di Yarega. Un requisito essenziale per tali operazioni *in situ* «modificate» è l'esistenza di uno strato roccioso strutturalmente compatto vicino al giacimento di petrolio, preferibilmente sotto il giacimento stesso, ma a non più di 30 metri di distanza. Il primo passo in un'operazione mineraria di questo genere sarebbe quello di scavare un pozzo nella roccia compatta sotto il giacimento; dal fondo del pozzo verrebbe estesa in senso orizzontale una serie di gallerie fino a formare sotto tutta la superficie del giacimento una sorta di rete stradale.

Un giacimento tipico di petrolio contiene, in ordine ascendente, acqua, petrolio e gas. Lavorando nelle gallerie sottostanti, i minatori trivellerebbero il giacimento verso l'alto; per tutta la zona d'acqua si adotterebbero rivestimenti

Il drenaggio per gravità è un sistema di coltivazione che dipende anch'esso dall'esistenza di una formazione di rocce compatte situate sotto il giacimento di petrolio. In questa roccia viene scavata una rete di gallerie *(a)* il più vicino possibile al fondo del giacimento. Nella roccia serbatoio vengono invece praticati per trivellazione numerosi fori di raccolta di breve lunghezza (dai 30 ai 45 metri). Attraverso una rete di condutture il petrolio fluisce, per effetto della forza di gravità, in un bacino di raccolta, da dove viene pompato alla superficie per la raffinazione. Il sistema unico di pompaggio rende economicamente vantaggiosa la coltivazione basata sul drenaggio per gravità.

Il drenaggio per mezzo di pompe è un sistema alternativo per sfruttare i depositi di petrolio che si sono svuotati in maniera incompleta, ma che non si trovano sopra una formazione di rocce compatte. Qui le gallerie di accesso vengono scavate sopra il deposito e molti fori di breve lunghezza vengono praticati nella roccia serbatoio verso il basso. Poiché ogni foro di trivellazione richiede una propria pompa, questo sistema è meno economico del drenaggio per gravità. Nell'URSS le miniere di petrolio di questo genere, che si servono di iniezioni di vapore per diminuire la viscosità, recuperano petrolio greggio pesante che non si può estrarre col metodo delle pompe sistemate alla superficie.

solidi, che però, nei punti in cui entrassero anche nella zona del petrolio, verrebbero perforati in modo da permettere alla forza di gravità di far defluire il petrolio in un sistema di raccolta allestito nelle gallerie sottostanti. Da qui il petrolio verrebbe pompato alla superficie attraverso il pozzo. Poiché i fori di trivellazione sarebbero corti (dai 30 ai 45 metri), sarebbe economicamente possibile trivellare verso l'alto il giacimento in punti molto ravvicinati, in modo da favorire al massimo il drenaggio. Con opportuni controlli, sarebbe possibile sfruttare il giacimento in maniera uniforme, in modo tale da garantire un'elevata percentuale di recupero del petrolio. Se necessario, si potrebbe ricorrere, come nel caso di Yarega, alla stimolazione per mezzo del vapore, in modo da assicurare il fluire del petrolio nei fori di trivellazione.

Uno dei nostri studi eseguiti su commissione, che prendeva a modello il giacimento di Wheeler Ridge nei pressi di Bakersfield (California), calcola che il costo del petrolio ottenuto da una tale combinazione di coltivazione in sotterraneo e di drenaggio per gravità sarebbe di 14 dollari il barile. (Questa cifra, espressa in dollari al valore del 1977, tiene conto anche delle royalty e di un certo margine di profitto.)

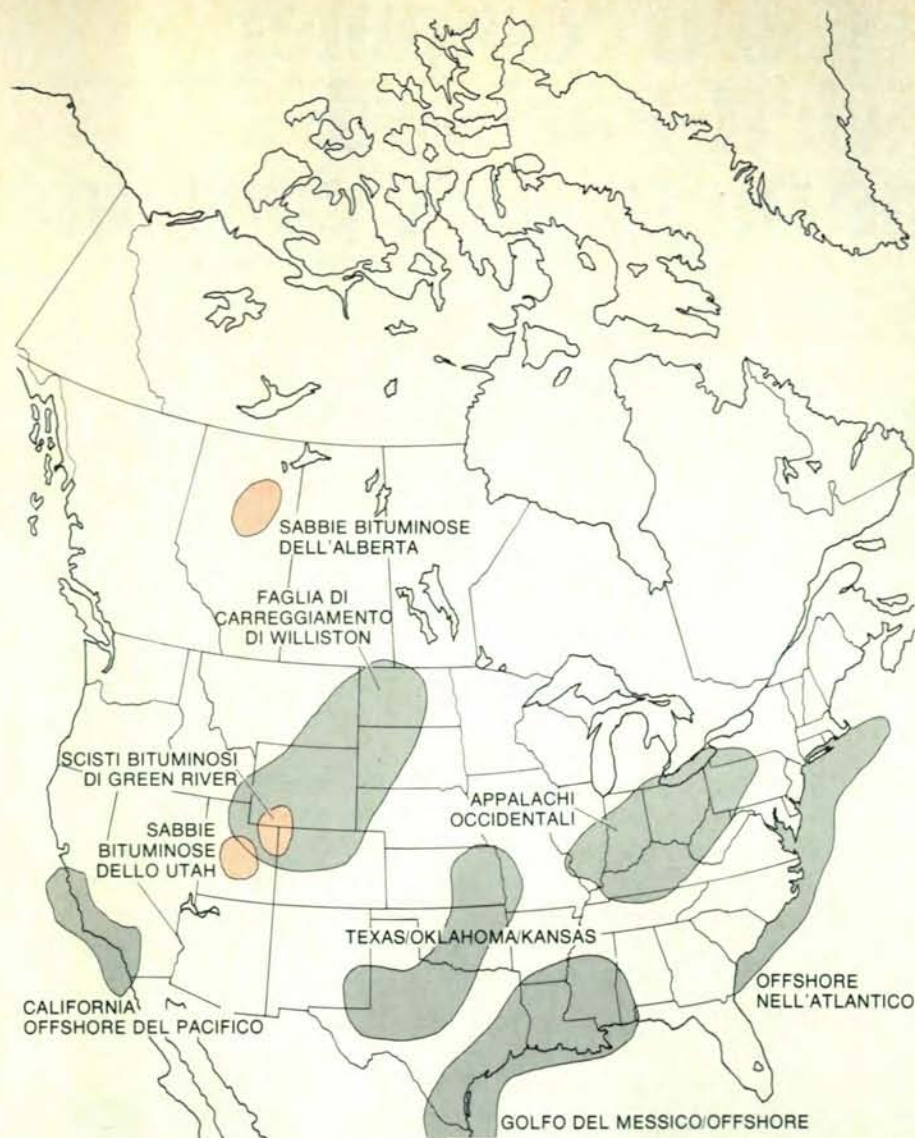
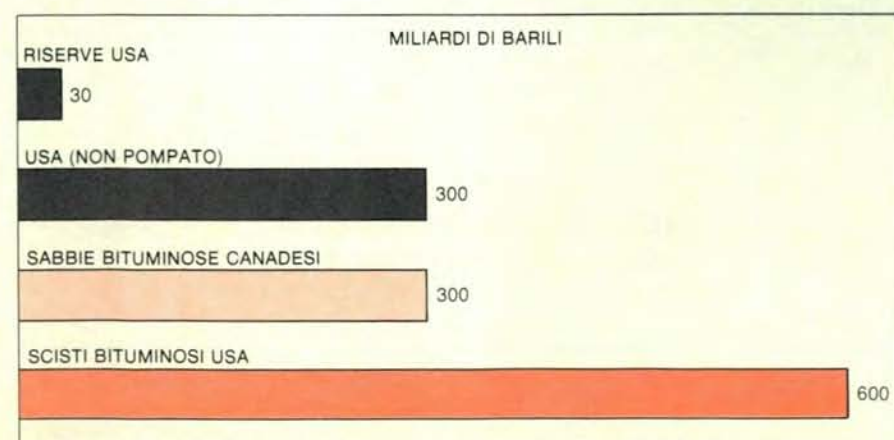
Nei sondaggi esplorativi si è soliti fermarsi quando il foro di trivellazione entra nello strato petrolifero. Di conseguenza nessuno sa quanti giacimenti di petrolio degli Stati Uniti siano sostenuti alla base da uno strato di roccia abbastanza compatto da permettere lo scavo del sistema di gallerie richiesto dalle miniere che sfruttano la gravità. Prima di poter fare una qualsiasi progettazione a proposito di queste miniere, sarebbe necessario un ambizioso programma di trivellazioni esplorative, in modo da stabilire con esattezza quali giacimenti americani si trovino sopra uno strato roccioso compatto sufficientemente vicino al fondo del giacimento stesso.

Già dalle normali pratiche di prospezione petrolifera risulta per contro che molti depositi di petrolio degli Stati Uniti si trovano sopra uno strato roccioso abbastanza compatto da permettere un sistema modificato di estrazione *in situ* molto più simile al metodo adottato a Yarega. In questo caso il sistema di gallerie verrebbe scavato sopra il giacimento e i fori di trivellazione sarebbero praticati verso il basso. Il drenaggio per gravità sarebbe sostituito dall'uso di una pompa per ogni foro di trivellazione. Questa combinazione di coltivazione e di pompaggio sarebbe necessariamente più complessa e più costosa del sistema di drenaggio per gravità, ma non è detto che lo sia in maniera addirittura proibitiva.

Per esempio, un altro dei nostri studi, che prendeva come modello il giacimento di Irma nell'Arkansas meridionale, calcola che il costo complessivo del petrolio di una miniera di questo genere sarebbe di 16 dollari il barile. (Anche questa stima e tutte le altre che seguono sono rapportate in dollari al valore del 1977.) Se si rendesse necessaria, come a Yarega, l'iniezione di vapore, il prezzo stimato salirebbe a 24 dollari il barile. Coi prezzi del 1980, che per il petrolio importato superano i 30 dollari il barile, l'estrazione per pompaggio incomincia ad apparire allettante anche dal punto di vista economico.

Pur essendo di primaria importanza, le considerazioni economiche costituiscono però in ogni caso soltanto uno dei tre fondamentali criteri di fattibilità per la coltivazione mineraria di un giacimento petrolifero. Gli altri due sono il grado di sicurezza dei minatori e l'entità dei danni arrecati all'ambiente naturale. Ovviamente non è accettabile nessun sistema di coltivazione che metta eccessivamente a repentaglio la sicurezza e la salute dei minatori o che degradi eccessivamente l'ambiente o che presenti entrambi questi inconvenienti. (L'avverbio «eccessivamente» in questo caso è appropriato perché non esiste nessun sistema di coltivazione che non esponga i minatori a qualche rischio o che non alteri in qualche misura l'ambiente.) Da questi due punti di vista, un sistema modificato di coltivazione *in situ* di un giacimento petrolifero merita una valutazione media se non addirittura ottima.

Per quel che riguarda la sicurezza dei minatori, qualunque lavoro sotterraneo in prossimità di depositi di idrocarburi comporta rischi associati agli incendi e alle esplosioni. In sistemi *in situ* modificati, tuttavia, la rete delle gallerie si troverebbe a una certa distanza dal giacimento vero e proprio e sarebbe contenuta nella roccia compatta. Inoltre, il petrolio estratto dal giacimento fluirebbe alla superficie in un sistema ermeticamente chiuso. Potenzialmente comunque esiste-

Le sabbie bituminose dell'Alberta sono sfruttate per mezzo di tecniche di coltivazione a giorno. Qui si vedono due fosse scavate in un deposito di sabbia bituminosa dalla Syncrude Canada Ltd. Le *dragline* accumulano la roccia bituminosa in andane: se ne possono vedere due in funzione dalla parte delle fosse più vicina a chi guarda. La roccia viene poi deposta su nastri trasportatori per mezzo di escavatori a ruota con tazze, di cui se ne può vedere uno all'opera proprio sopra la *dragline* a sinistra. I lunghi nastri trasportatori portano la roccia fino all'impianto di trasformazione (non visibile nella fotografia), dove il bitume viene convertito in petrolio greggio sintetico.

Cinque fra le principali regioni produttrici di petrolio dei 40 stati contigui degli USA sono indicate qui insieme con la zona atlantica *offshore* non ancora produttiva. Sono presentate *(in colore)* anche tre risorse non tradizionali di idrocarburi, le sabbie bituminose dell'Alberta e dello Utah e gli scisti bituminosi della formazione di Green River nello Utah, nel Colorado e nel Wyoming. Ognuna delle cinque regioni ha giacimenti petroliferi esauriti, dove forse si potrebbe però estrarre petrolio con tecniche di coltivazione mineraria. La regione più promettente è la California, dove, oltre ai normali depositi di petrolio, un vasto strato di sedimenti diatomitici contiene idrocarburi.

Raffronto tra le riserve stimate di petrolio degli USA, circa 30 miliardi di barili, e la quantità stimata di petrolio che rimane attualmente o che rimarrà nei serbatoi esauriti: 300 miliardi di barili. Questa cifra è pari alla quantità totale stimata di petrolio sintetico disponibile nelle sabbie bituminose dell'Alberta *(in colore chiaro)* e a circa metà della quantità recuperabile dagli scisti bituminosi di Green River, che possono produrre più di 15 galloni di petrolio per tonnellata *(in colore scuro)*.

rebbe sempre il rischio di una fuga di gas. Pertanto tali miniere richiederebbero certamente norme di sicurezza estremamente rigorose, come quelle che vengono imposte attualmente alle miniere di carbone in sotterraneo.

Quanto all'impatto sull'ambiente, la coltivazione in sotterraneo in genere ha un notevole potenziale di degradazione sia in termini di materiale sterile della miniera, sia in termini di subsidenza geologica (lo sprofondamento del terreno conseguente all'attività estrattiva). Il sistema modificato di coltivazione *in situ* di un giacimento di petrolio presenta una minaccia minima all'ambiente da tutti e due i punti di vista. La quantità di materiale scavato portato alla superficie e successivamente eliminato è piccola rispetto a quella ottenuta con i sistemi di estrazione tradizionali. Anche le possibilità di subsidenza sono irrilevanti. La maggior parte del materiale estratto da una miniera di questo genere è costituita dal petrolio stesso. Acqua e roccia, per esempio, sarebbero prodotte in piccole quantità e l'acqua potrebbe venire iniettata di nuovo nel giacimento, minimizzando ulteriormente le possibilità di subsidenza. Ricapitolando, la coltivazione modificata *in situ* di un giacimento petrolifero costituirebbe un processo notevolmente pulito.

Negli Stati Uniti la maggior parte dell'estrazione di minerali è attuata non in sotterraneo ma con una di queste tre tecniche di coltivazione a giorno: coltivazione a gradini *(strip mining)*, coltivazione a fossa *(open-pit mining)* e coltivazione a scarpata *(terrace mining)*. Queste stesse tecniche si possono applicare alla coltivazione mineraria di un deposito di petrolio. Tutte e tre hanno una caratteristica in comune: la necessità di rimuovere terra e roccia su scala molto vasta. Ciò richiede notevoli investimenti in attrezzature pesanti, se si vuole che i costi unitari rimangano economici.

Proprio come i sistemi *in situ* modificati sono appropriati per lo sfruttamento di giacimenti petroliferi situati circa 150 metri al di sotto della superficie, così le tecniche di coltivazione a giorno sono applicabili ai depositi meno profondi. Descriveremo i tre metodi nell'ordine. La coltivazione a gradini, una tecnica usata frequentemente nei bacini carboniferi delle regioni medio-occidentali e occidentali degli Stati Uniti, ha inizio con la rimozione del materiale di copertura mediante lo scavo di una fossa lunga e stretta effettuato con una *dragline* o un escavatore a cucchiaio. Di solito la fossa è larga da 15 a 30 metri e può essere lunga più di un chilometro e mezzo e profonda quel tanto che occorre per mettere a nudo il deposito del minerale. Il minerale viene quindi rimosso e trasportato in un impianto di trasformazione. In seguito viene scavata una seconda fossa parallela alla prima. Il materiale di copertura di questa fossa viene scaricato nella prima, dopo di che si procede a rimuovere il minerale. La sequenza si

protrae ripetutamente via via che si susseguono le operazioni estrattive.

La coltivazione a gradini è un sistema efficiente per i depositi che si trovano a una profondità massima di 50-60 metri in terreni generalmente pianeggianti. Il recupero del terreno soggetto a questo tipo di estrazione comporta operazioni relativamente semplici, quali lo spianamento dei mucchi di materiale di copertura, la sostituzione dello strato di superficie e la sistemazione di nuove piante. In molti siti delle regioni medio-occidentali i terreni sfruttati con la coltivazione a gradini vengono in questo modo restituiti abitualmente all'agricoltura.

La coltivazione a imbuto è adatta ai depositi di minerali caratterizzati da una topografia superficiale irregolare e anche a quelli che si trovano a una profondità di oltre 55-60 metri. Tanto il materiale di copertura quanto il deposito di minerale vengono rimossi con grandi escavatori a pala e portati fuori dello scavo mediante un sistema di strade di carreggio. È possibile che vengano usati anche sistemi di nastri trasportatori. Il carico di minerale è avviato all'impianto di trasformazione, mentre il materiale di copertura viene scaricato a una certa distanza dallo scavo. In generale risulta impossibile servirsi del materiale di copertura per riempire lo scavo.

La terza tecnica di coltivazione a giorno, la coltivazione a scarpata, è una variazione della coltivazione a imbuto alla quale si fa ricorso quando il deposito di minerale copre una superficie estesa, ma è relativamente poco profondo. La si può considerare una specie di coltivazione a gradini che si avvolge a spirale verso l'esterno, ma in questo caso il materiale di copertura viene portato via per mezzo di autocarri e accantonato almeno temporaneamente invece di essere riscaricato subito nello scavo. Dato però che la coltivazione a scarpata racchiude alla fine una vastissima superficie di terreno esaurito, di solito è di uso pratico servirsi di una parte del terreno di copertura come materiale da ripiena.

Secondo alcune stime, negli Stati Uniti la produzione di petrolio mediante tecniche di coltivazione a giorno è economicamente possibile. Uno studio, che prendeva come modello il giacimento di Kern River in California e come tecnica la coltivazione a scarpata, avanza l'ipotesi che il costo complessivo del petrolio sarebbe di 11 dollari il barile. Un altro studio, che prendeva come modello il giacimento di Edna in California e come tecnica la coltivazione a imbuto, prevede un costo di 15 dollari il barile.

Con il giacimento petrolifero di Santa Cruz in California come modello e adottando la coltivazione a gradini come tecnica, il costo previsto è di 18 dollari il barile. Quando fu ipotizzato un sistema modificato di coltivazione a scarpata per il deposito di sabbia bituminosa nei pressi di Sunnyside, nello Utah, il costo stimato si aggirava intorno ai 21 dollari il barile, una cifra solo leggermente più elevata di quella del petrolio greggio sintetico che viene ricavato dalle sabbie bituminose canadesi.

Sebbene la coltivazione a giorno sia un po' più sicura di quella *in situ* modificata, il suo influsso sull'ambiente è maggiore. Nelle vicinanze delle miniere ci si può attendere gas, polvere ed esalazioni tossiche. Tanto il materiale di copertura quanto gli scarti degli impianti di trasformazione presenterebbero notevoli problemi di smaltimento. Ovviamente è possibile, a un certo prezzo, ridurre la portata di queste conseguenze sfavorevoli e arrivare anche a eliminarle completamente, ma quando ci si chiede quali siano esattamente le restrizioni ambientali da porre alla coltivazione a giorno di un giacimento di petrolio, la domanda per il momento rimane senza valida risposta. Intanto il Bureau of Mines sta svolgendo una ricerca intesa a identificare e a quantificare questi problemi ambientali e a proporre possibili soluzioni.

Perché attualmente negli Stati Uniti non si estrae petrolio mediante tecniche minerarie? Una risposta, che era valida in passato, era che la nazione poteva andare avanti benissimo senza bisogno di ricorrere a tali tecniche per lo sfruttamento dei giacimenti petroliferi. A mano a mano però che i prezzi del petrolio salgono, questa risposta diventa sempre meno logica. A dire il vero, sono stati già avviati esperimenti pilota di coltivazione mineraria di depositi di petrolio, compresa anche l'estrazione di una certa quantità di petrolio da scisti bituminosi. Un consorzio formato da 17 società, fra cui alcune delle più importanti compagnie petrolifere statunitensi, ha portato a termine di recente un'azione dimostrativa, durata 30 mesi, sulla possibilità pratica di ricavare petrolio greggio sintetico dalla formazione di scisti bituminosi nei pressi di Rifle (Colorado). Verso la fine del 1978, alla conclusione del periodo di dimostrazione, la distillazione degli scisti bituminosi aveva prodotto circa 100 000 barili di petrolio greggio alla media di 25 galloni di greggio per ogni tonnellata di scisto. In un altro sito del Colorado è attualmente in corso una ricerca che viene condotta su scisti bituminosi che sono stati frantumati con esplosivi.

Intanto un produttore di petrolio sta progettando di aprire una miniera a imbuto per sfruttare gli idrocarburi presenti in una formazione di sedimenti di diatomee profonda circa 120 metri che si trova nel campo McKittrick in California. Un altro produttore, che lavora in cooperazione con il Dipartimento dell'energia, sta indagando la possibilità di recuperare idrocarburi da una formazione di sabbia bituminosa nello Utah, servendosi di metodi di estrazione *in situ*. Sembra legittimo prevedere che in futuro si assisterà a uno sfruttamento accelerato, mediante coltivazione mineraria, sia dei depositi non tradizionali di idrocarburi degli Stati Uniti, sia delle notevoli risorse di petrolio ancora presenti nei giacimenti che sono già stati sfruttati con la tradizionale tecnica del pompaggio.

Struttura e funzione del cromosoma eucariota

Negli organismi eucarioti i cromosomi sono strutture assai complesse, ma duttili e dinamiche; queste qualità hanno consentito loro di svolgere un importante ruolo nell'evoluzione contribuendo alla variabilità genetica

di Lino C. Polito e Maria Furia

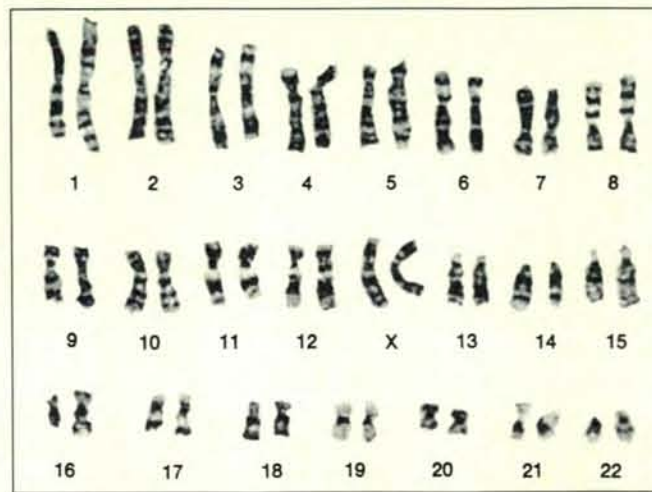
I cromosomi sono le strutture cellulari portatrici dell'informazione genetica in tutti gli organismi; vi è però una netta differenza fra la loro organizzazione negli organismi procarioti e quella tipica degli eucarioti. Nei primi il cromosoma appare costituito da una lunga molecola di DNA praticamente nudo, che non è distinguibile al microscopio ottico e viene quindi studiato principalmente al microscopio elettronico o mediante analisi genetica. Negli eucarioti invece i cromosomi si presentano come strutture estremamente complesse di cui il DNA è solo uno dei componenti e sono generalmente presenti in coppie di elementi uguali fra loro (omologhi), il cui numero è caratteristico per ogni organismo. I cromosomi eucarioti possono essere osservati anche al microscopio ottico, ma tale osservazione è limitata al momento della divisione cellulare sia mitotica che meiotica; nelle altre fasi del ciclo cellulare infatti non è visibile alcuna struttura cromosomica ben definita, anche se il DNA resta associato agli altri componenti cromosomici, principalmente proteine, in una sostanza nucleare diffusa che prende il nome di cromatina.

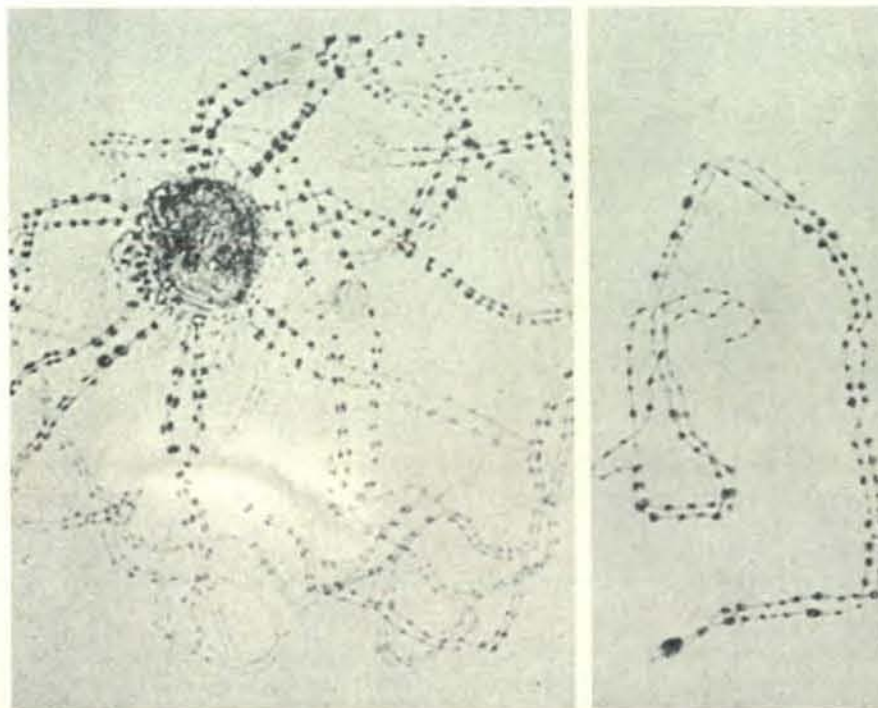
Solo all'inizio del processo di divisione cellulare la cromatina inizia a condensarsi in strutture più compatte che sono quindi più adatte al processo di segregazione e più facilmente visibili. Il processo di condensazione della cromatina è progressivo e raggiunge il suo apice durante la metafase: a questo stadio della divisione cellulare ogni cromosoma ha una sua caratteristica morfologia, data dalla forma, dalla lunghezza e dalla posizione del centromero e delle costrizioni secondarie. La disposizione di tutti i cromosomi sul piano equatoriale rende poi questa fase la più adatta e più agevole per lo studio del cariotipo. Il materiale genetico delle cellule eucariote attraversa quindi nelle successive divisioni cellulari un processo alterno di condensazione e dispersione che è strettamente correlato alla sua attività funzionale: lo stato diffuso della cromatina appare infatti un requisito indispensabile allo svolgersi dei processi di replicazione e trascrizione.

Durante l'interfase sono facilmente distinguibili due differenti gradi di organizzazione del materiale genetico: infatti, mentre la maggior parte si presenta diffusa (eucromatina), una parte residua rimane allo stato condensato e si colora più intensamente con coloranti specifici del DNA (eterocromatina). Tratti eterocromatici sono distinguibili anche su preparati di cromosomi metafasici, dove l'eterocromatina ha una distribuzione caratteristica: generalmente si notano lungo un cromosoma larghe regioni eucromatiche accanto a piccole regioni eterocromatiche, localizzate quasi sempre alle estremità cromosomiche (telomeri) e ai centromeri. Questa distribuzione è tipica della eterocromatina detta costitutiva in quanto costituisce una costante morfologica del cromosoma; è possibile poi distinguere un secondo tipo di eterocromatina che viene detta facoltativa in quanto rappresenta uno stato transitorio del cromosoma. Zone di eterocromatina di questo tipo non sono infatti necessariamente presenti in tutte le cellule di un organismo, né in tutte le fasi di differenziamento di un dato tipo cellulare: un esempio di eterocromatina facoltativa è fornito nelle femmine dei mammiferi dalla coppia di cromosomi sessuali (o cromosomi X) uno dei quali, fin dai primi stadi dello sviluppo, diviene eterocromatico nelle cellule somatiche costituendo il cosiddetto corpo di Barr *(si vedano le fotografie nella pagina a fronte e in alto a pagina 132)*. Nei mammiferi marsupiali, il cromosoma X che diviene eterocromatico è sempre di origine paterna, mentre nei mammiferi placentali questo processo è del tutto casuale; infatti in ogni cellula somatica già nei primi stadi di sviluppo embrionale, uno dei due cromosomi X diviene eterocromatico e conserva tale stato in tutte le cellule figlie. Il soma femminile nei mammiferi placentali è quindi costituito da un mosaico di cellule in cui alternativamente è stato reso eterocromatico il cromosoma X di origine materna o paterna.

Il diverso grado di organizzazione strutturale dell'eucromatina e dell'eterocromatina riflette anche un diverso grado di funzionalità; l'incorporazione di precursori radioattivi dell'RNA seguita con tecniche autoradiografiche porta a concludere che l'eterocromatina è molto meno attiva dell'eucromatina nei processi di trascrizione e sembra perciò corrispondere a zone di scarsa (se non nulla) funzionalità genetica.

Allo stadio della profase, sia in mitosi che in meiosi, sono distinguibili al microscopio ottico lungo l'asse del cromosoma zone larghe che si alternano a zone sottili; le zone larghe, che presentano spesso una forma tondeggiante, costituiscono le più piccole subunità cromosomiche riconoscibili al microscopio ottico e sono chiamate cromomeri *(si vedano le immagini in basso a pagina 132)*. Il numero, le dimensioni e la posizione dei cromomeri su ogni cromosoma sono caratteristici di un determinato tessuto a un dato stadio dello sviluppo. Per esempio, se si osserva lo stesso cromosoma nella profase mitotica e in quella meiotica si nota che il numero dei cromomeri mitotici è sensibilmente più basso. Il numero, la forma e le dimensioni dei cromomeri variano con continuità anche nei vari stadi della profase meiotica: nel leptotene sono generalmente piccoli e numerosi mentre nel tardo pachitene considerevolmente più grandi e di numero minore. A ogni stadio l'intera struttura del cromosoma è quindi riorganizzata e ciascun cromomero è ricostruito in una configurazione stadio-specifica. Tra cellule germinali di sesso diverso sono riscontrabili differenze nel numero dei cromomeri anche allo stesso stadio della profase meiotica; per esempio negli oociti e negli spermatociti del grillo *Acheta* si osservano rispettivamente 178 e 227 cromomeri per l'intero genoma. In speciali tipi di cromosomi, i cromosomi *lampbrush* o piumosi (detti anche da alcuni autori a spazzola) e in quelli politenici, i cromomeri sono stati identificati come siti suscettibili di attiva trascrizione *(si vedano le due figure a pagina 133)*. I cromosomi piumosi hanno dimensioni insolitamente elevate e morfologia particolare, caratterizzata dalla presenza di anse di varia lunghezza che emergono dai cromomeri e conferiscono loro il tipico aspetto piumoso. Cromosomi di questo tipo possono essere osservati per esempio durante la spermatogenesi di *Drosophila*, ma soprattutto durante l'oogenesi di molti vertebrati, in particolar modo negli oociti giganti degli anfibi. In questi organismi la profase meiotica si protrae per un periodo di tempo considerevolmente lungo, dovuto principalmente a un allungamento dello stadio di diplotene. In questo stadio i cromoso-

L'analisi cromosomica o cariotipo può essere facilmente eseguita durante la metafase mitotica. Per individuare con maggiore sicurezza i cromosomi sono state introdotte tecniche di colorazione e digestione enzimatica che consentono di distinguere su di essi caratteristiche zone chiare e scure. Nella fotografia a sinistra è riprodotto un cariotipo di donna normale come appare nella piastra metafasica. A destra i cromosomi omologhi sono stati disposti in coppie numerate consentendo di riconoscere le 22 paia di autosomi e il paio di cromosomi sessuali (XX).

Sono qui rappresentate schematicamente le principali fasi della mitosi e della meiosi. Mentre il primo tipo di divisione cellulare ha lo scopo di dare origine a due cellule figlie aventi lo stesso corredo cromosomico della cellula da cui derivano, il secondo tipo di divisione produce quattro cellule aventi corredo cromosomico dimezzato (gameti) che possono, fondendosi con altri gameti, produrre uno zigote avente corredo cromosomico diploide. Va notato che l'appaiamento dei cromosomi omologhi presente nella meiosi è assente nella mitosi e che i centromeri dei cromatidi fratelli si separano nella meiosi solo nella seconda divisione (divisione equazionale).

Nel nucleo interfasico di due cellule ottenute dalla mucosa boccale di donna *(fotografie in alto)* è possibile riconoscere il caratteristico corpo di Barr formato da cromatina sessuale. Nella corrispondente cellula di uomo *(fotografia in basso)* il corpo di Barr è invece assente.

Stadi di pachitene o diplotene precoce in microsporociti di *Trillium* che mostrano come durante la meiosi avvenga l'appaiamento tra omologhi e come si dispongano i cromomeri. Il corpo scuro nella microfotografia a sinistra è il nucleolo; nella figura a destra sono visibili parecchi chiasmi.

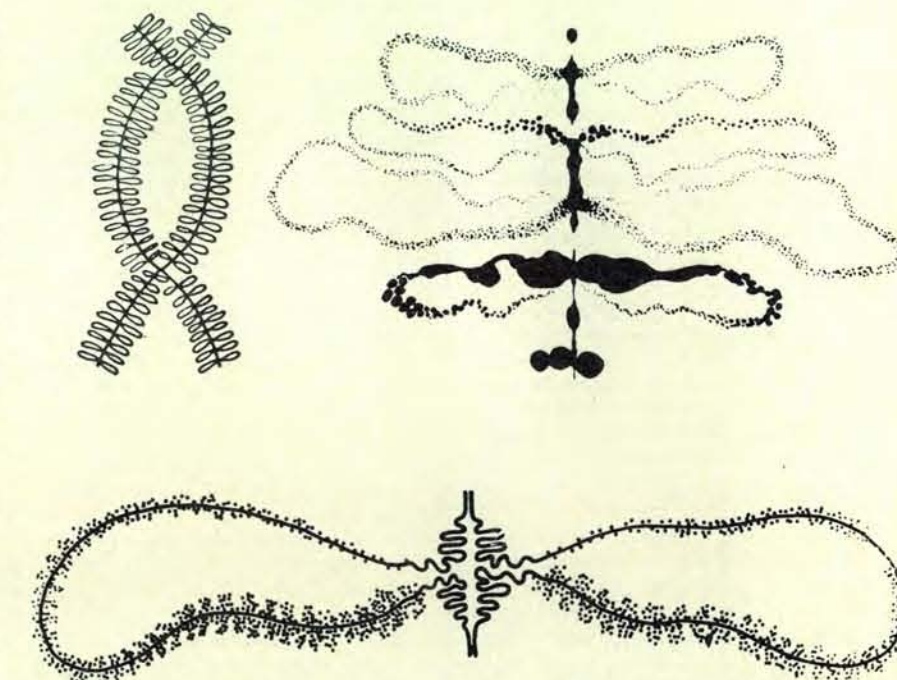
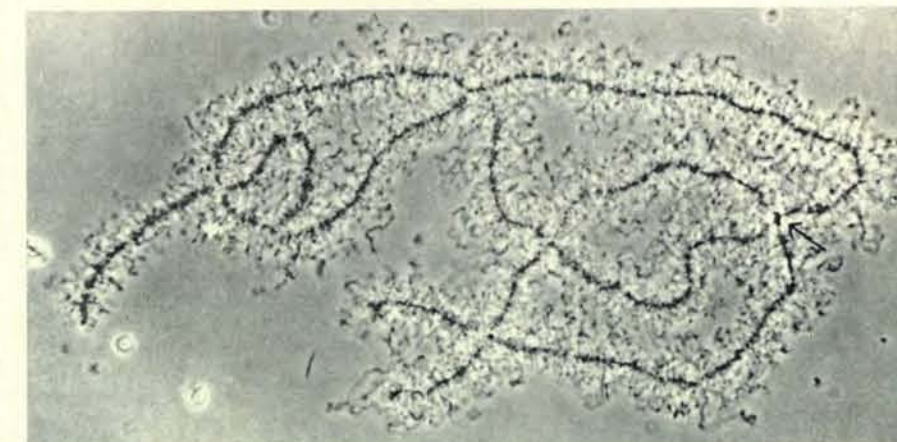
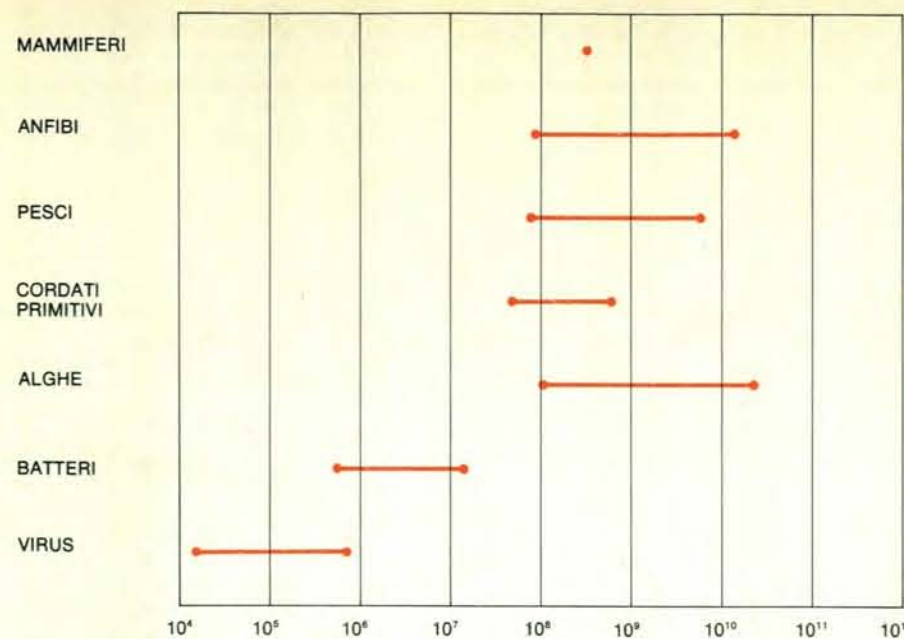
mi piumosi raggiungono lunghezze pari a 500-1000 micrometri e solo nei successivi stadi meiotici ritornano alle abituali dimensioni di 50-100 micrometri. A questa mancata condensazione è correlata una intensa sintesi di RNA e proteine che porta a un considerevole aumento di volume della massa citoplasmatica; il fenomeno è infatti tipico di uova che attraversano un lungo periodo di accrescimento durante il quale vengono accumulate le sostanze di riserva che costituiscono il tuorlo. H. G. Callan e H. C. MacGregor, digerendo con DNAsi preparati di cromosomi piumosi, hanno dimostrato che le anse sono costituite da DNA e che vi è continuità fra il filamento di DNA svolto nelle anse e quello fortemente spiralizzato presente nel cromomero.

I cromosomi politenici, chiamati anche cromosomi giganti, derivano da cicli multipli di replicazione cui non fa seguito la divisione cellulare. I filamenti di DNA così prodotti (cromatidi) rimangono quindi appaiati e formano strutture di dimensioni estremamente elevate. I cromosomi politenici sono stati studiati soprattutto nei ditteri, dove sono presenti in alcuni tessuti specializzati delle larve, quali le ghiandole salivari, ma sono diffusi in molte altre specie di insetti e in una varietà di altri organismi, tra cui le piante e i protozoi, dove appaiono per esempio durante lo sviluppo del macronucleo nei ciliati ipotrichi. Il numero dei cicli di endoreplicazione da cui originano i cromosomi giganti può essere calcolato paragonando il contenuto di DNA dei nuclei politenici con quello dei normali nuclei diploidi. È stato così stabilito che nelle cellule delle ghiandole salivari delle larve di *Drosophila melanogaster* si succedono ben 9 cicli di endoreplicazione, dal momento che le cellule politeniche presentano, in luogo del normale contenuto diploide, un contenuto di DNA pari a 1024 volte quello del corredo cromosomico della cellula germinale (corredo aploide).

Tratti diversi di un cromosoma possono essere replicati con differente molteplicità durante il processo di politenizzazione. Sempre in *Drosophila*, per esempio, mentre la maggior parte del materiale cromosomico, tra cui principalmente le zone eucromatiche, è soggetto a 9 cicli di replicazione, le zone eterocromatiche sono sottoreplicate o non vengono replicate affatto; altre regioni poi, come quelle che contengono i geni per l'RNA ribosomale, sono replicate a un livello intermedio fra quello dell'eucromatina e dell'eterocromatina.

Lungo l'asse dei cromosomi politenici è visibile, al microscopio ottico, l'alternarsi di zone più dense, chiamate bande, a zone meno dense, chiamate interbande; la differente densità rispecchia una diversa concentrazione di DNA *(si veda la figura in alto a pagina 135)*. È ipotesi corrente che le bande dei cromosomi politenici siano formate dalla giustapposizione laterale dei cromomeri presenti in ciascun cromatide. La posizione, le dimensioni e la morfologia delle bande nei cromosomi politenici sono caratteristiche sufficientemente costanti in differenti tessuti della stessa specie; alcune bande però si decondensano a precisi stadi dello sviluppo della larva o in seguito a stimoli ambientali: la decondensazione, chiamata anche *puffing*, corrisponde a uno svolgimento locale delle fibre cromosomiche ed è correlata a una intensa sintesi di RNA *(si veda la figura in basso a pagina 135)*. A uno stesso stadio di sviluppo tessuti diversi mostrano differenti schemi di decondensazione, mentre in uno stesso tessuto lo schema è caratteristico e varia nel corso dello sviluppo.

L'esistenza di schemi specifici ha suggerito che il *puff* sia la manifestazione citologica dell'espressione genica e, corrispondentemente, la banda costituisca la sede dell'unità genica. Questa idea è in accordo con i primi studi sui cromosomi politenici in cui fu notato che la distribuzione e la sequenza delle bande e delle interbande sono specifiche per ogni cromosoma. Riarrangiamenti del materiale genetico quali duplicazioni, delezioni, inversioni e traslocazioni che variavano la posizione dei marcatori nella mappa genetica alteravano contemporaneamente l'ordine e la posizione delle bande sui cromosomi politenici; è stato quindi possibile correlare la disposizione lineare dei geni nella mappa genetica con la disposizione lineare delle bande nella mappa citologica. Nelle ghiandole salivari di *Drosophila melanogaster*, C. B. Bridges ha contato al microscopio ottico approssimativamente 5059 bande; questo dato concorda fondamentalmente con le determinazioni più recenti e più precise effettuate con l'uso del microscopio elettronico e con il numero di geni ipotizzabili per quest'organismo. Burke Judd con i suoi collaboratori, allo scopo di definire con precisione la relazione esistente fra cromomeri e geni, ha effettuato un'analisi genetica approfondita della regione 3A2-3C2 del cromosoma X di *Drosophila melanogaster.*

Il primo passo di questo esperimento fu l'isolamento di mutanti rappresentativi di ogni gene presente nella regione. I mutanti puntiformi isolati erano poi incrociati a due a due per valutare la possibilità di integrazione funzionale tra le varie mutazioni. Infatti, quando due mutazioni recessive vengono introdotte nello stesso organismo e si ottiene uno sviluppo normale, si ritiene che appartengano a unità funzionali differenti; se viceversa la combinazione delle due mutazioni continua a presentare un fenotipo mutato, si assume che le mutazioni interessino la stessa unità funzionale. Questo tipo di analisi genetica prende il nome di test di complementazione. Fu visto così che le mutazioni puntiformi si raggruppavano in 14 gruppi funzionali, che aggiunti ai loci *zeste* e *white* che delimitano rispettivamente a sinistra e a destra la regione 3A2-3C2 davano per l'intervallo esaminato un totale di 16 gruppi di complementazione. Questo numero coincide con quello delle bande distinguibili al microscopio elettronico per la stessa regione. Mediante l'analisi citologica di varie delezioni fu possibile assegnare i gruppi di complementazione alle bande della regione esaminata e il risultato fu che ogni banda corrispondeva a un gruppo di complementazione. La conclusione tratta da Judd e collaboratori fu che essenzialmente in ciascuna banda viene espressa solamente una funzione genica. A conclusioni analoghe sono giunte, sempre in *Drosophila*, le ricerche di G. Levefre e B. Hochman rispettivamente sul cromosoma X e sul cromosoma 4; tali conclusioni però si basano esclusivamente sui dati forniti da mutazioni letali o che comunque portano a gravi alterazioni fenotipiche, mentre è noto che non tutti i geni mutati portano a letalità o a effetti fenotipici ben visibili. Le difficoltà esistenti nell'individuare loci genetici le cui mutazioni non modificano né il normale sviluppo né il normale fenotipo suggeriscono perciò che l'identificazione della banda con l'unità genica possa implicare una semplificazione eccessiva. Inoltre, tenendo presente che nella regione studiata da Judd ogni banda contiene mediamente 25 000 coppie di basi, e che 1000 coppie sono sufficienti a codificare per una proteina di dimensioni medie, ogni banda nella regione *zeste-white* contiene materiale genetico sufficiente a codificare 25 proteine di dimensioni medie e quindi, se la correlazione fra banda e unità è corretta, nel cromomero è sempre presente un grande eccesso di acido dessossiribonucleico.

La microfotografia in alto mostra due cromosomi piumosi o a spazzola (in inglese *lampbrush*) ottenuti da oocita di *Triturus*. I due omologhi sono tenuti insieme da numerosi chiasmi e dal cromocentro indicato dalla freccia. L'aspetto piumoso è dovuto alla presenza di anse laterali. I disegni in basso rappresentano dei modelli schematici di questi stessi cromosomi. In alto a sinistra sono rappresentati cromosomi omologhi uniti da due chiasmi; un ingrandimento dell'asse cromosomico *(a destra)* mostra sui due cromatidi fratelli due anse di morfologia identica. In basso appaiono, notevolmente ingrandite, due anse su cui si stanno sintetizzando catene di RNA.

Quantità di DNA per nucleo aploide in organismi che possiedono differente complessità evolutiva. Le linee in colore indicano le quantità di DNA minime e massime per ogni gruppo di organismi; come si può notare tutti i mammiferi possiedono all'incirca la medesima quantità di DNA.

È noto da tempo che negli eucarioti la quantità totale di DNA presente nel genoma aploide, detta anche valore C, è spesso in largo eccesso rispetto alla quantità di DNA necessaria per il numero di geni calcolato sulla base del numero delle funzioni ipotizzabili in un dato organismo. Per esempio, nell'uomo il corredo cromosomico aploide contiene 3×10^9 coppie di basi di DNA e quindi 3×10^6 geni potenziali, mentre il numero di geni ipotizzato arriva al massimo a valori compresi tra 10^4 e 10^5; questo eccesso di DNA presente negli eucarioti suggerisce la possibilità che solo una parte di esso sia geneticamente funzionale. S. Ohno ha calcolato che se tutto il DNA dell'uomo fosse portatore di informazione genetica, poiché il tasso di mutazione è comunemente di 10^{-5}/locus/generazione, in ogni gamete dovrebbero accumularsi ben 30 nuove mutazioni; poiché è lecito aspettarsi che una parte di queste mutazioni risulti letale o comunque fortemente svantaggiosa, questo carico mutazionale avrebbe dovuto portare già da un pezzo all'estinzione dell'uomo e di ogni altro organismo superiore. Sulla base di queste considerazioni Ohno ha concluso che nell'uomo meno del 10 per cento del DNA deve essere geneticamente attivo.

Anche se, salendo lungo la scala evolutiva, all'aumentare della complessità degli organismi corrisponde quasi sempre un aumento della quantità di DNA *(si veda la figura in questa pagina)*, non esiste una proporzionalità diretta fra la quantità di DNA presente in un dato organismo e la sua complessità genetica. Infatti, all'interno di molte famiglie di piante e di animali il contenuto di DNA varia ampiamente anche tra specie strettamente correlate la cui affinità farebbe supporre la presenza di un numero di geni sostanzialmente simile; gli anfibi, per esempio, hanno un contenuto di DNA che oscilla addirittura fra 1 e 100×10^9 coppie di basi e per molti di essi l'ammontare di DNA è maggiore di quello della specie umana, senz'altro più complessa dal punto di vista evolutivo. La mancanza di proporzionalità fra la complessità di un organismo e il suo contenuto di DNA è nota come il paradosso del valore C e porta ancora una volta a suggerire che la maggior parte del DNA potrebbe non svolgere funzione genetica o almeno non portare istruzione per proteine. Queste considerazioni hanno trovato supporto nei risultati ottenuti dallo studio chimico-fisico del DNA degli eucarioti, basato principalmente su tecniche di ultracentrifugazione in gradienti di densità di sali di cesio e di denaturazione e rinaturazione del DNA e, in tempi più recenti, sull'uso di enzimi di restrizione e della clonazione di DNA ricombinante. Si è così evidenziato nel DNA degli eucarioti l'esistenza di tre diversi tipi di sequenze: un primo tipo rappresenta generalmente più del 50 per cento del DNA, in alcuni casi fino a circa l'80 per cento, ed è costituito da sequenze presenti una sola volta nel genoma aploide e perciò dette uniche; un secondo tipo, che costituisce circa il 15 per cento del DNA totale, comprende sequenze ripetute nel genoma da 10^3 a 10^5 volte e chiamate mediamente ripetute; la frazione minore comprende, infine, circa il 10 per cento del totale e consiste di sequenze molto corte (da poche coppie di basi a non più di un centinaio), che si ripetono nel genoma fino a 10^6 volte e sono pertanto dette altamente ripetute. Le sequenze altamente ripetute costituiscono spesso il cosiddetto DNA satellite perché, avendo una composizione in basi generalmente diversa da quella del DNA totale, hanno anche una diversa densità e pertanto formano in gradienti di sali di cesio una o più bande distinte da quella del restante DNA.

La tecnica dell'ibridazione citologica in situ, messa a punto da Mary Lou Pardue e Joe Gall ha permesso di dimostrare che nei cromosomi mitotici di numerosi organismi il DNA satellite è localizzato principalmente nelle zone dell'eterocromatina centromerica *(si veda la figura in basso a pagina 137)*. Questo dato non implica necessariamente che le sequenze altamente ripetute siano contigue in uno stesso tratto del DNA, perché la spiralizzazione della fibra cromatinica (cromonema) potrebbe portare ad affiancarsi nella stessa zona del cromosoma sequenze linearmente lontane. Esperimenti analoghi dimostrano che non vi è una localizzazione cromosomica preferenziale per le sequenze mediamente ripetute, che risultano invece diffuse in tutte le regioni del cromosoma.

Risultati basati sulle cinetiche di rinaturazione di frammenti di DNA di varia lunghezza e sull'osservazione al microscopio elettronico di DNA rinaturato indicano che le sequenze singole sono intercalate con le sequenze mediamente ripetute: l'esistenza di questo tipo di organizzazione in gran parte del genoma è stata dapprima osservata in *Xenopus*, in *Drosophila* e nel riccio di mare e confermata via via in altri organismi, benché vi sia variazione nella lunghezza relativa e nella distribuzione reciproca dei due tipi di sequenze.

Molti autori hanno cercato di legare in una visione unitaria la organizzazione sul DNA delle sequenze singole e ripetute agli interrogativi posti dal paradosso del valore C: per esempio, Charles A. Thomas, Jr., riprendendo una vecchia idea di Callan e L. Lloyd, aveva suggerito che ogni locus genetico potesse essere costituito da una serie di geni identici e che il variare della quantità di DNA fra specie affini derivasse principalmente dalla variazione della molteplicità di queste sequenze, cioè dalla loro ridondanza nel genoma, che veniva così visto come costituito da blocchi di geni multipli. Questo modello è però in contrasto con il dato che le sequenze singole costituiscono la frazione di gran lunga più abbondante in ogni genoma e che a questa frazione sono associati i geni strutturali, cioè i geni che portano l'informazione per le proteine. Dati di ibridazione fra l'RNA presente nel nucleo e le diverse frazioni del DNA hanno infatti dimostrato che, anche se l'RNA nucleare viene trascritto dalle sequenze uniche e da quelle mediamente ripetute, nel citoplasma gli RNA che portano le istruzioni per le proteine (RNA messaggeri) ibridano esclusivamente con il DNA a sequenza singola, tranne alcune eccezioni, quali per esempio l'RNA messaggero per gli istoni che ibrida con la frazione mediamente ripetuta.

L'esistenza di famiglie di geni multipli, quali i geni per gli istoni, per le immunoglobuline, per gli RNA di trasporto, per l'RNA ribosomale e altri, rappresenta l'eccezione e non la norma, indicando comunque che in alcuni tratti di cromosoma geni identici sono sequenzialmente ripetuti, secondo il modello suggerito da Thomas. La grossa differenza nella quantità di DNA di specie affini non sembra dovuta alla variazione del numero dei geni strutturali, associati principalmente alla frazione a sequenza unica, quanto alle variazioni della frazione del DNA mediamente ripetuta, probabilmente deputata a funzioni regolative di struttura e funzione.

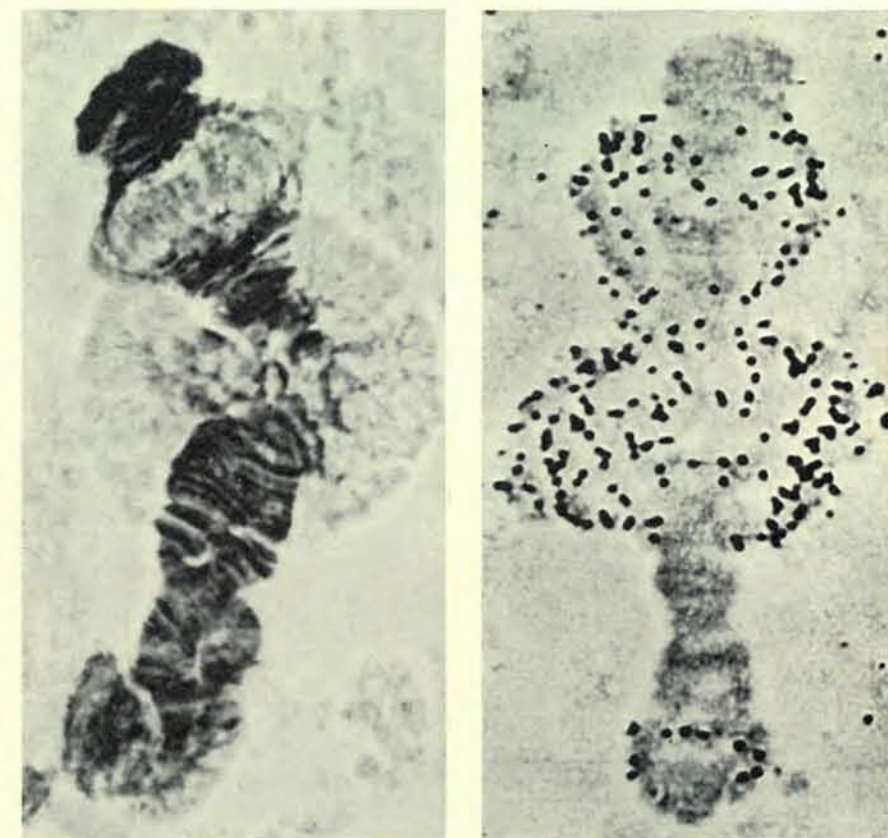
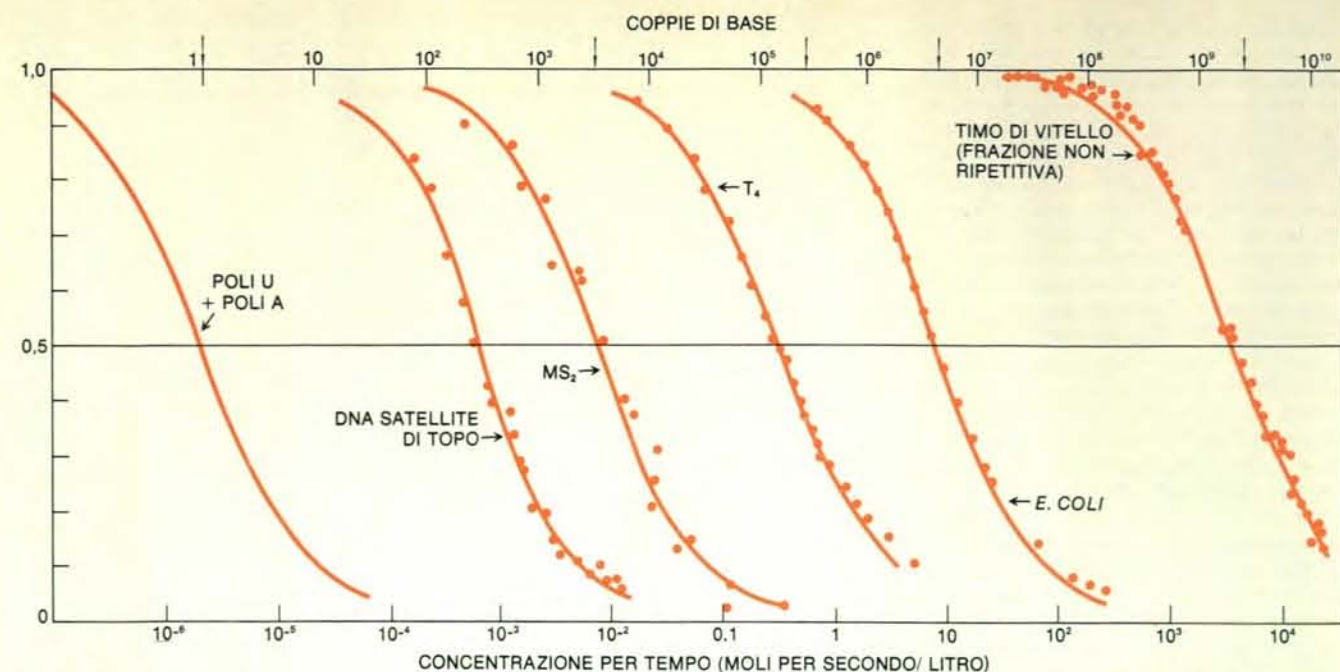
Recentemente l'uso delle tecniche di ingegneria genetica e di clonazione del DNA ricombinante ha reso possibile la purificazione di numerosi geni. L'osservazione al microscopio elettronico delle strutture DNA/RNA, formate per ibridazione di RNA messaggeri purificati alle sequenze pure del DNA omologo, ha permesso di osservare che anche all'interno di un gene possono essere presenti sequenze che sono trascritte, ma non sono poi presenti nell'RNA messaggero e non vengono quindi tradotte. La funzione svolta da questo tipo di sequenze non è ancora nota, anche se è logico supporre che la loro esistenza debba necessariamente avere delle implicazioni rilevanti nella espressione dell'attività genica.

Sequenze del tipo descritto prendono il nome di inserzioni e la loro presenza è stata riportata in numerosi geni, quali quelli delle globine in varie specie di mammiferi, per l'ovalbumina nel pollo, per le immunoglobuline nel topo, per la fibra della seta nel baco da seta e in alcuni geni strutturali di virus animali, oltre che nei geni non propriamente strutturali per l'RNA di trasporto (tRNA) nel lievito e per quello ribosomale (rRNA) in *Drosophila*.

Come si può notare i geni strutturali citati codificano tutti per prodotti specializzati di cellule altamente differenziate. È principalmente per questo tipo di geni che è attualmente disponibile una dettagliata analisi molecolare; è necessaria l'analisi della struttura di numerosi altri geni per stabilire quanto la presenza delle inserzioni sia diffusa nell'organizzazione genica degli eucarioti.

Il numero e le dimensioni delle inserzioni finora descritte variano da caso a caso: ve ne sono per esempio solo due nei geni di alcune catene delle immunoglobuline, ma non meno di sette nel gene dell'ovoalbumina; la loro lunghezza poi è variabile da meno di 100 a più di 1000 coppie di basi. È interessante notare che nel gene dell'ovoalbumina la lunghezza totale delle inserzioni supera di almeno tre volte la parte effettivamente tradotta. Se la presenza delle inserzioni dovesse dimostrarsi un fenomeno di portata generale, l'estrapolazione di questo dato porterebbe a concludere che in un gene eucariota possano esservi più sequenze *nonsense* che sequenze *sense;* è evidente che questo dato porterebbe nuovi elementi alla nostra comprensione del paradosso del valore C.

Fin dagli anni cinquanta attraverso lo studio dei cromosomi al microscopio elettronico si è tentato di correlare le va-

Nella microfotografia sono chiaramente visibili cromosomi politenici ottenuti da cellule di ghiandole salivari di *Drosophila melanogaster*. Gli omologhi di ogni coppia si presentano strettamente appaiati; su ogni cromosoma è evidente una caratteristica distribuzione di bande e interbande.

La microfotografia a sinistra rappresenta il cromosoma gigante n. 4 di *Chironomus tentans* con regioni rigonfie dette *puff*. Il puff è sede di attiva trascrizione dell'RNA come dimostra l'immagine autoradiografica a destra ottenuta dopo l'incorporazione di precursori radioattivi dell'RNA.

Sono qui rappresentate le cinetiche di rinaturazione di campioni di DNA di varia provenienza. Per la misurazione si considera il tempo impiegato da una data concentrazione di DNA a elica singola per riformare molecole a doppia elica. I grafici riportano la percentuale relativa di doppia elica presente in funzione della concentrazione iniziale del DNA e del tempo in cui viene effettuata la misura. La ricostituzione della doppia elica richiede la collisione delle due catene complementari e pertanto la velocità del processo è proporzionale alla probabilità che un'elica incontri quella complementare. A parità di concentrazione e in condizioni sperimentali controllate, il tempo necessario per la rinaturazione aumenta con l'aumentare delle dimensioni del genoma. Come si vede la rinaturazione nel batterio *E. coli* è più lenta di quella nei fagi. Sequenze ripetute quali poli U/poli A e quelle del DNA satellite di topo rinaturano più velocemente del DNA a sequenza singola.

riazioni morfologiche visibili al microscopio ottico durante le varie fasi del ciclo cellulare a eventuali variazioni nell'organizzazione a livello molecolare. L'osservazione di cromosomi metafasici al microscopio elettronico ha permesso di visualizzare un insieme di fibre, di diametro variabile fra 23 e 30 nanometri, avvolte in maniera irregolare all'interno della struttura cromosomica e che ai telomeri non presentano estremità terminali, così che ogni cromatide sembra costituito da una singola fibra avvolta su se stessa. Preparati di cromatina interfasica mostrano la presenza di fibre analoghe, benché di diametro medio leggermente inferiore (23 nanometri); il diametro di ciascuna fibra comunque non è costante, ma varia con regolarità lungo il suo asse, ricordando l'aspetto di grani allineati su di un filo *(si veda la figura in basso a pagina 139)*.

Esperimenti di digestione delle fibre mediante tripsina o pronasi portano a una notevole riduzione del loro diametro rivelando l'esistenza di una parte interna resistente alle proteasi ma sensibile alle DNAsi, il cui diametro, di 2,4 nanometri, avvalora l'ipotesi che essa sia costituita da una singola molecola di DNA. L'insieme di queste osservazioni ha permesso a E. J. Du Praw di formulare un modello per la struttura cromosomica secondo cui ogni cromatide è costituito da una unica lunga molecola di DNA che viene mantenuta in una configurazione super-spiralizzata dall'associazione con molecole proteiche, formando così la fibra cromatinica visibile al microscopio elettronico. Tale ipotesi è confermata da dati sperimentali quali quelli di Ruth Kavenoff e Bruno H. Zimm che, misurando le proprietà viscoelastiche del DNA da cromosomi metafasici di *Drosophila,* hanno dimostrato l'esistenza di molecole di DNA di lunghezza corrispondente a quella prevista per il contenuto di DNA di ogni singolo cromosoma; esperimenti analoghi condotti da altri autori sui cromosomi di lievito hanno portato ugualmente a concludere che il filamento di DNA contenuto in ogni cromatide è una molecola continua e priva di interruzioni.

Du Praw e G. F. Bahr hanno valutato che la lunghezza della molecola di DNA contenuta in un cromosoma è 100 volte maggiore della lunghezza della fibra cromatinica e questa a sua volta è 50 volte più lunga della dimensione raggiunta dal cromosoma stesso in metafase (5-10 micrometri). Un grado di condensazione così elevato non può essere raggiunto da una semplice spiralizzazione del filamento di DNA ed è indispensabile supporre l'esistenza di avvolgimenti di ordine superiore. Le componenti proteiche della cromatina svolgono probabilmente un ruolo essenziale nel determinare sia la condensazione del DNA all'interno delle fibre, sia l'organizzazione delle fibre all'interno del cromosoma ed è di notevole interesse chiarire quali siano le interazioni molecolari che si stabiliscono fra le varie proteine cromatiniche e fra queste e il DNA.

Le componenti proteiche della cromatina sono distinguibili essenzialmente in proteine basiche o istoniche e proteine acide o non-istoniche. Gli istoni si suddividono in 5 gruppi principali (H1, H2A, H2B, H3, H4) che differiscono tra di loro per il contenuto relativo di amminoacidi basici quali lisina e arginina *(si veda la tabella al centro a pagina 138)*.

Il rapporto esistente nella cromatina fra il contenuto in peso di DNA e quello degli istoni è di 1:1 in ogni cellula eucariota. Inoltre le sequenze amminoacidiche di alcuni tipi di istoni (soprattutto H4 e H3, ma anche H2A e H2B) sono molto simili anche in organismi evolutivamente distanti e ciò suggerisce l'idea che gli istoni partecipino a interazioni così specifiche da essersi mantenuti praticamente inalterati nel corso dell'evoluzione. In contrasto con ciò, tuttavia, gli istoni mostrano una varietà di modificazioni chimiche postsintetiche quali metilazioni, acetilazioni e fosforilazioni che potrebbero svolgere un ruolo rilevante in una loro interazione differenziale con il DNA. Il ruolo degli istoni nell'organizzazione strutturale del cromosoma è stato in gran parte chiarito da esperimenti di digestione della cromatina con varie DNAsi e di ricostituzione in vitro a partire da istoni e DNA. Questi esperimenti portano a stabilire nella cromatina l'esistenza di una unità basica identificabile in tutte le cellule eucariote in un ottamero formato da due molecole di ognuno degli istoni H3, H4, H2A, H2B intorno a cui si avvolge un tratto di DNA di lunghezza compresa fra 160 e 200 coppie di basi *(si veda la figura in alto a pagina 139)*. Strutture di questo tipo sono connesse da tratti di DNA di lunghezza variabile e al loro interno il DNA è notevolmente condensato: infatti il diametro di un nucleosoma è di circa 10

Quando nel DNA in esame sono presenti sequenze singole e sequenze ripetute, la cinetica di rinaturazione seguirà l'andamento della curva in colore pieno che è tipico del DNA degli eucarioti. I tratti che rinaturano con velocità maggiore del DNA di *E. coli (curva in colore chiaro)* sono dovuti alla presenza di sequenze altamente e mediamente ripetute. Come dimostrano anche i grafici nella pagina a fronte, l'analisi della cinetica di rinaturazione permette di risalire alle percentuali con le quali i differenti tipi di sequenze sono presenti nel genoma.

nanometri, mentre la lunghezza prevista per un filamento di DNA di 200 paia di basi è di 68 nanometri. Esperimenti di ricostituzione *in vitro* indicano anche che l'istone H1 non è necessario per ottenere le figure di diffrazione ai raggi X tipici della cromatina nativa. Il ruolo svolto da questo istone nella struttura della cromatina non è ancora chiaro, anche se alcune indicazioni suggeriscono che esso, interagendo con il DNA nucleosomico e con quello che separa nucleosomi adiacenti, regoli la lunghezza del tratto internucleosomico determinando così lo stato finale di condensazione della cromatina.

Le figure di diffrazione dei raggi X della cromatina e le osservazioni al microscopio elettronico, hanno suggerito ad alcuni autori un modello secondo cui il filamento dei nucleosomi si avvolge ulteriormente su se stesso formando una struttura di tipo solenoidale con un passo di 11 nanometri. Il diametro previsto per un avvolgimento di questo tipo, di circa 23 nanometri, si dimostrerebbe in accordo con il diametro descritto per la fibra cromatinica.

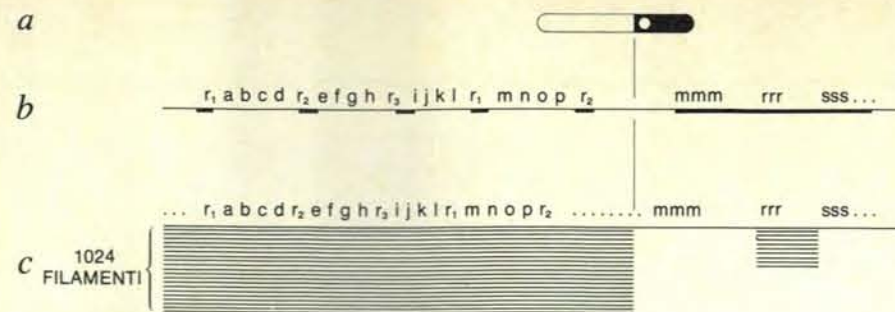
Per quanto riguarda le proteine non istoniche, va notato che questa classe di proteine è essenzialmente eterogenea e la sua composizione relativa varia nei diversi tipi cellulari; inoltre cellule differenziate impegnate in attiva sintesi hanno livelli di proteine non istoniche più alti di quelli di cellule indifferenziate o differenziate ma sinteticamente inattive. Queste osservazioni hanno suggerito l'ipotesi che le proteine non istoniche svolgano un ruolo regolativo, determinando la porzione del genoma che viene trascritta all'interno di una data cellula a ogni fase dello sviluppo.

Benché il meccanismo con cui queste proteine eserciterebbero il loro controllo sulla trascrizione non sia noto, è stato proposto un modello in cui viene postulato che la regolazione della interazione delle proteine non istoniche con il DNA avvenga tramite la loro fosforilazione e defosforilazione; modificazioni chimiche postsintetiche, con conseguenti alterazioni dei profili elettroforetici delle proteine non istoniche, sono infatti spesso correlate a variazioni nella fisiologia o nello sviluppo di una cellula.

Risultati recenti indicano inoltre che alcune proteine non istoniche possono essere responsabili del mantenimento della morfologia nei cromosomi metafasici: infatti, se da un preparato di cromosomi si rimuovono con tecniche opportune gli istoni e la maggior parte delle proteine non istoniche, il DNA resta ugualmente organizzato in strutture compatte la cui morfologia rimane identica a quella dei cromosomi intatti. Un ulteriore trattamento di questo preparato con tripsina ne distrugge la morfologia specifica, mentre la digestione del DNA non ne altera la struttura. L'analisi della composizione dello scheletro proteico, che sembra quindi responsabile del mantenimento della morfologia, ha rivelato la presenza di circa 30 proteine appartenenti alla classe delle proteine non istoniche.

In conclusione, mentre le proteine istoniche sembrano avere principalmente una funzione strutturale nell'organizzazione della cromatina, vi sono forti indicazioni che la classe delle proteine non istoniche possa svolgere un ruolo essenziale nei processi che determinano sia la morfologia del cromosoma sia la sua funzionalità.

I complessi meccanismi che sono alla base dei processi di differenziamento cellulare fanno supporre l'esistenza negli

Ibridazione in situ di un preparato di cromosomi mitotici di topo nei quali il DNA satellite è stato marcato radioattivamente. La marcatura è caratteristicamente localizzata ai centromeri e ai telomeri, indicando in queste regioni la sede cromosomica delle sequenze satelliti.

Modello dell'organizzazione strutturale del cromosoma metafasico e di quello politenico. In *a* è rappresentato un cromosoma metafasico; il tratto scuro indica la regione eterocromatica. In *b* è schematizzato il filamento uninemico del DNA nel cromosoma metafasico: r_1, r_2, r_3 indicano sequenze mediamente ripetute intercalate a sequenze singole (a-d; e-h; i-l; m-p), mmm e sss sequenze altamente ripetute e rrr sequenze mediamente ripetute adiacenti, quali i geni ribosomali. In *c* è rappresentata la replicazione differenziale delle varie sequenze durante la politenizzazione.

DESCRIZIONE	NOMENCLATURE ALTERNATIVE			PESO MOLECOLARE
RICCHI DI LISINA	I	f1	H1	21 000
MODERATAMENTE RICCHI DI LISINA	IIb1	f2b	H2B	13 700
	IIb2	f2a2	H2A	14 500
RICCHI DI ARGININA	III	f3	H3	15 300
	IV	f2a1	H4	11 300

In questa tabella sono elencati le nomenclature usate nei diversi laboratori e i pesi molecolari degli istoni (proteine della cromatina a basso peso molecolare) ottenuti dal timo di vitello.

eucarioti di numerosi e articolati sistemi di regolazione dell'attività genica.

La complessità degli organismi superiori rende però difficile in molti casi persino l'identificazione e la mappatura dei geni strutturali; e inoltre, anche in sistemi nei quali l'analisi genetica è più agevole, come per esempio in *Drosophila*, è spesso difficile correlare ai loci genetici già individuati e mappati i prodotti genici a essi associati. L'identificazione di sistemi di regolazione genica si presenta negli eucarioti estremamente complessa; infatti in questi ultimi non è stato ancora identificato nessun gene o sito avente una funzione chiaramente regolativa. Un aspetto generale dell'organizzazione funzionale del genoma eucariota è che i geni che concorrono all'espressione di una stessa funzione non sono contigui sul DNA, come avviene spesso nei procarioti, ma possono comparire anche su cromosomi diversi.

Nei procarioti la contiguità dei geni che concorrono all'espressione di uno stesso fenotipo ha permesso l'instaurarsi di speciali sistemi di regolazione. È ben noto infatti che geni contigui possono essere trascritti in una unica molecola di RNA messaggero, sotto il controllo di siti di regolazione anch'essi adiacenti ai geni strutturali; l'insieme funzionale dei geni strutturali e dei siti di regolazione è noto come operone *(si veda la figura in alto a pagina 140)*.

La differente organizzazione esistente negli eucarioti porta a concludere che il modello di regolazione di tipo operone poligenico è difficilmente estendibile a questi organismi; per spiegare l'espressione simultanea di geni funzionalmente correlati bisogna quindi ipotizzare la presenza di meccanismi di controllo diversi e probabilmente più complessi. Sembrerebbe, quindi, che negli organismi superiori ogni unità genica debba corrispondere a un'unità di trascrizione. Tuttavia la taglia dei trascritti nucleari, i cosiddetti RNA giganti, è molto più grande del valore atteso per trascritti corrispondenti a proteine di dimensioni medie (~ 300 000 dalton per un trascritto di mille basi), raggiungendo a volte un peso molecolare di svariati milioni, mentre le molecole di RNA messaggero presenti nel citoplasma mostrano le dimensioni attese. Si deve supporre che i trascritti nucleari subiscano nel passaggio dal nucleo al citoplasma un processo di maturazione post-trascrizionale che ne riduce le dimensioni; in tal modo alcune sequenze potrebbero essere eliminate dai primi prodotti di trascrizione e solo una parte del trascritto iniziale passerebbe al citoplasma per essere successivamente tradotta. Processi di maturazione di questo tipo sono già noti per i prodotti di trascrizione di alcuni geni, tra cui quelli che presentano inserzioni, ma è ancora troppo presto per valutare quanto diffusi siano tali meccanismi.

Nel quadro della regolazione dell'attività genica negli eucarioti, meritano un cenno particolare i casi in cui una attivazione selettiva viene ottenuta attraverso processi di amplificazione genica. L'amplificazione genica può essere definita come la replicazione differenziale di una parte del genoma. Fenomeni di questo tipo si verificano negli insetti *Rhynchosciara, Drosophila* e *Acheta* e riguardano principalmente la regolazione del numero dei geni per l'RNA ribosomale. Il caso più noto e studiato in dettaglio è però quello che si verifica negli oociti degli anfibi. In questi animali le cellule somatiche presentano due nucleoli ognuno dei quali porta circa 450 geni per l'RNA ribosomale, mentre gli oociti già nei primi stadi dell'oogenesi presentano all'interno del nucleo un migliaio di nucleoli e un numero dei geni per l'RNA ribosomale amplificato di circa 2000 volte.

Fotografia al microscopio elettronico di un cromosoma umano allo stadio di metafase. Entrambi i cromatidi presentano fibre del diametro di circa 25 nanometri. L'autore della fotografia è A. Baué.

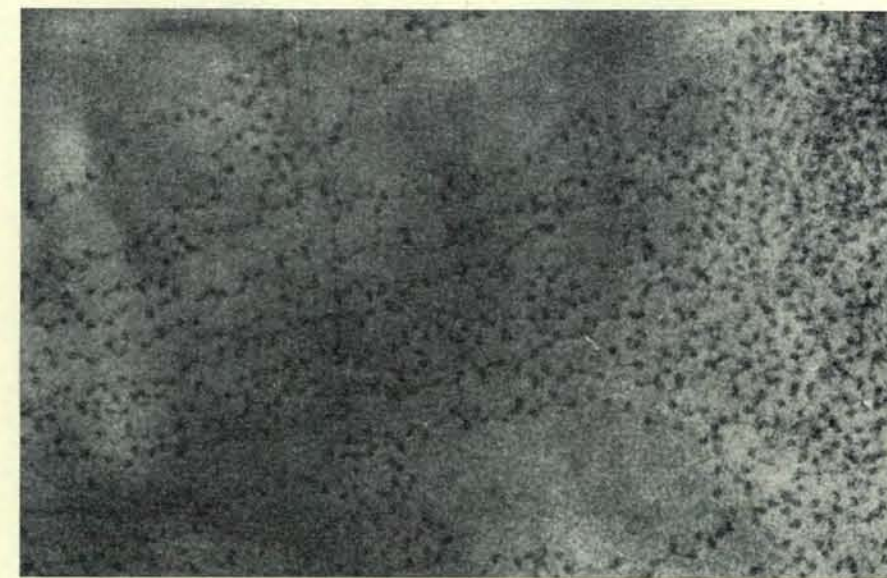
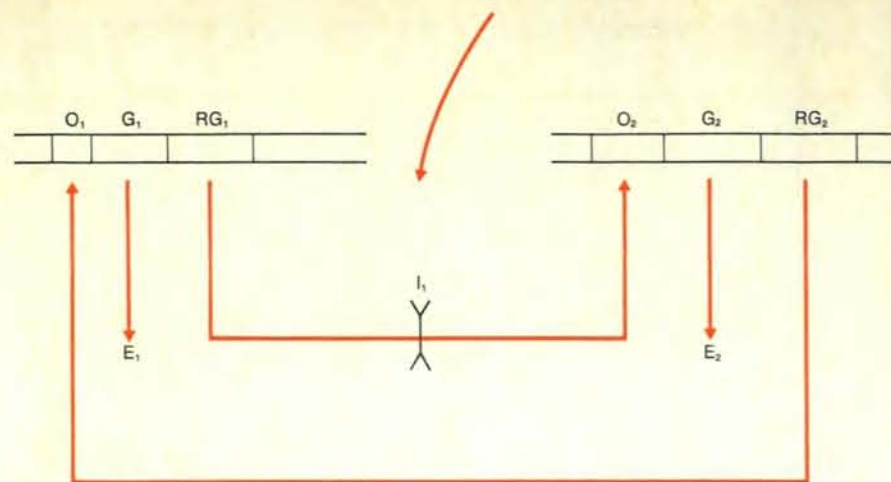
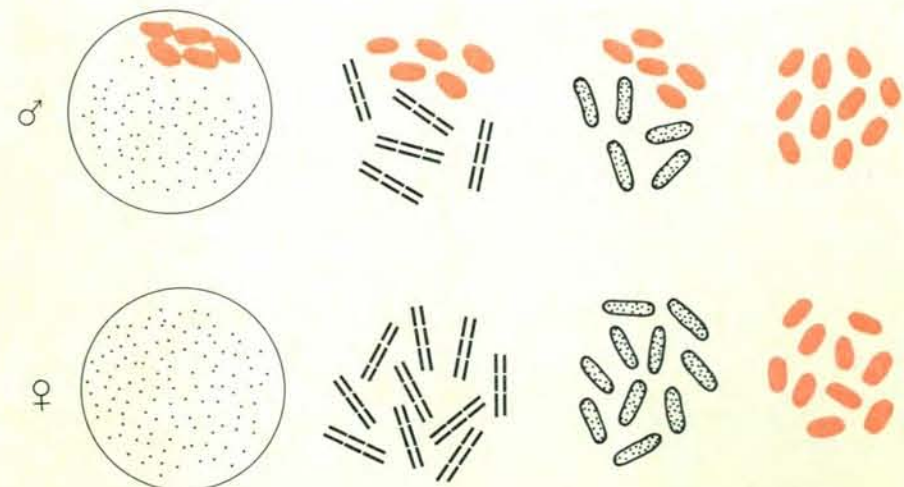
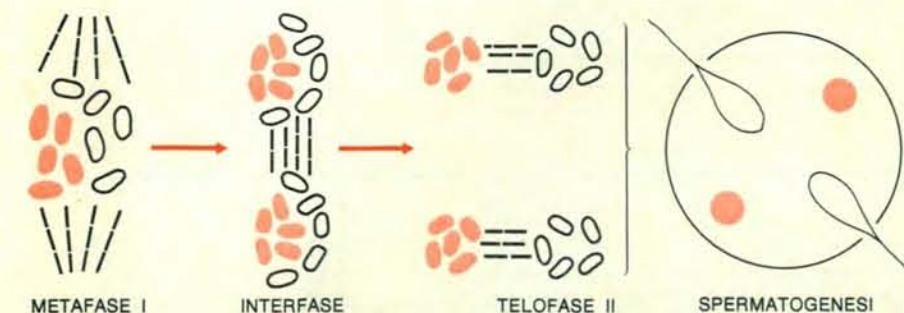
Recentemente è stato descritto da Robert T. Schimke e dai suoi collaboratori un fenomeno di amplificazione che riguarda l'acquisizione di resistenza alla ametopterina (nota in commercio come Metotrexate) da parte di alcuni cloni cellulari di topo. Il Metotrexate è un analogo dell'acido folico e inibisce la diidrofolatoreduttasi, un enzima che svolge un ruolo indispensabile nella sintesi delle purine. Poiché le purine sono essenziali nella replicazione cellulare, il Metotrexate è largamente usato come citostatico nella terapia di alcuni tumori, anche se spesso le cellule acquisiscono resistenza al farmaco attraverso varie vie, quali l'alterazione della permeabilità di membrana, dell'affinità dell'enzima per la ametopterina e infine l'aumento della quantità di enzima presente nelle cellule. Il gruppo di ricercatori di Stanford ha dimostrato che in alcuni casi l'acquisizione di resistenza può, invece, avvenire mediante un'amplificazione selettiva del gene per la diidrofolatoreduttasi; diviene così sempre più chiaro che l'espressione genica può essere modulata, oltre che per azione differenziale dei geni, anche attraverso variazioni del numero delle copie del gene stesso.

Una funzione rilevante nella regolazione dell'espressione genica è probabilmente svolta dai cosiddetti elementi genetici trasponibili, estesamente studiati nei procarioti e recentemente descritti in *Drosophila melanogaster* e nel lievito *Saccaromyces cerevisiae* (si veda l'articolo *Elementi genetici trasponibili* di S. N. Cohen e J. A. Shapiro in «Le Scienze», n. 140, aprile 1980). In realtà l'esistenza e il comportamento di tali elementi erano già stati postulati e descritti quarant'anni fa da Barbara McClintock nel mais, ma solo lo sviluppo delle tecniche di ingegneria genetica ha permesso di chiarirne la struttura molecolare e di delinearne la funzione. Gli elementi trasponibili sono segmenti di DNA capaci di spostarsi da un sito all'altro dei differenti cromosomi. In *Drosophila melanogaster* per esempio sono state descritte tre famiglie di geni, chiamate 412, copia e 297, che appartengono alla classe del DNA mediamente ripetitivo, rappresentano circa l'1 per cento del DNA totale e sono variabili in numero e localizzazione all'interno del genoma. L'idea corrente è che l'inserimento e il disinserimento di elementi trasponibili in siti specifici possa determinare l'attivazione o l'inattivazione dell'espressione dei geni situati in prossimità dei punti di inserzione.

Oltre a questa funzione regolativa l'introduzione di segmenti di DNA in siti di sequenza non omologa ha probabilmente svolto e svolge un ruolo importante nell'evoluzione, contribuendo ad aumentare la variabilità genetica.

È ben nota negli eucarioti l'esistenza di un ulteriore livello di regolazione mediante il quale l'inattivazione di tratti di cromosoma, di cromosomi interi e talvolta di interi assetti cromosomici si attua attraverso un processo di eterocromatizzazione legato allo sviluppo.

L'esempio più noto di regolazione dell'espressione genica mediante eterocromatizzazione è rappresentato dal fenomeno già citato dell'inattivazione di uno dei due cromosomi X presenti nelle femmine dei mammiferi; mediante questo meccanismo la quantità dei prodotti dei geni legati a questo cromosoma viene mantenuta uguale nei due sessi. Anche fra gli insetti vi sono casi di inattivazione di interi cromosomi. In molte specie di cocci-

Schema dell'organizzazione della fibra cromatinica. Un tetramero formato da due molecole degli istoni H3 e H4 si lega a un segmento di DNA e successivamente forma un legame stabile con due dimeri costituiti da una molecola dell'istone H2A e da una dell'istone H2B; infine si lega l'istone H1.

Fotografia al microscopio elettronico di un preparato di fibre cromatiniche ottenute da globuli rossi di pollo. Queste fibre sono ben riconoscibili per il tipico aspetto di grani allineati su un filo.

Modello di un circuito di regolazione di tipo operone proposto da F. Jacob e J. Monod. Il prodotto del gene regolatore RG_1 blocca, legandosi al sito operatore O_2, la trascrizione dell'operone 2, che contiene geni strutturali (G_2) e un secondo gene regolatore (RG_2). Il prodotto RG_2 regola a sua volta l'espressione di un secondo operone (operone 1), contenente i geni strutturali G_1 e lo stesso RG_1. L'azione regolatrice della proteina RG_1 è modulata dalla presenza della molecola I_1. Il complesso proteina $RG_1 + I_1$ non è più in grado di bloccare la trascrizione dell'operone 2; viene così sintetizzata la proteina RG_2 che, legandosi al sito O_1, blocca la trascrizione dell'operone 1.

Schema della divisione mitotica in maschi e femmine della pulce della farina. Nei due sessi è presente lo stesso numero di cromosomi; i cromosomi della femmina *(in basso nella figura)* presentano un comportamento normale durante le varie fasi dello sviluppo embrionale, mentre metà di quelli del maschio diventano eterocromatici in uno stadio precoce dello sviluppo. I cromosomi eterocromatici si raggruppano nell'interfase in una struttura densa chiamata cromocentro. Durante la profase essi appaiono notevolmente più condensati, mentre nello stadio della metafase tutti i cromosomi di entrambi i sessi presentano il medesimo grado di condensazione.

Schema della spermatogenesi nella pulce della farina. I cromosomi eucromatici e quelli eterocromatici si dividono equazionalmente durante la prima divisione meiotica; nella seconda divisione poi i due gruppi si separano e negli spermi maturi si ritrovano solo i cromosomi eucromatici.

di un intero assetto cromosomico aploide diviene eterocromatico nei primi stadi dell'embriogenesi e rimane eterocromatico nello sviluppo successivo. Nella pulce della farina, in cui questo processo è stato studiato in dettaglio, tutti i cromosomi appaiono eucromatici negli embrioni di entrambi i sessi, fino allo stadio di blastula. Successivamente negli individui di sesso maschile l'intero assetto cromosomico di origine paterna diviene eterocromatico; in questi maschi anche lo svolgimento della meiosi è anomalo. La prima divisione meiotica è infatti equazionale; l'assetto eucromatico e quello eterocromatico si duplicano e si dividono come in una divisione somatica. Nella seconda divisione meiotica invece l'assetto eucromatico e quello eterocromatico migrano ai poli opposti e solo i prodotti meiotici che contengono il corredo eucromatico formeranno spermi vitali *(si veda la figura in basso in questa pagina)*. Nei maschi della pulce della farina quindi solo il corredo cromosomico materno è geneticamente attivo e viene trasmesso alla progenie attraverso gli spermi; inoltre le meiosi della linea germinale maschile non presentano appaiamento degli omologhi e di conseguenza sono prive di ricombinazione genetica.

Sempre negli insetti, il genere *Sciara* presenta un complesso meccanismo di eliminazione selettiva di cromosomi anziché di semplice inattivazione; la perdita selettiva di un cromosoma potrebbe rappresentare un caso estremo di uno stesso meccanismo regolativo.

In *Sciara* gli zigoti di entrambi i sessi sono geneticamente identici; sia nel maschio che nella femmina, infatti, sono presenti tre cromosomi sessuali, due di origine paterna e uno di origine materna. La determinazione del sesso in questi organismi è collegata a un elaborato processo di eliminazione di cromosomi. La perdita di un solo cromosoma X di origine paterna porta alla formazione di un soma femminile di costituzione genetica XX, mentre quella di entrambi i cromosomi sessuali di origine paterna porta alla formazione di uno maschile di costituzione XO. A questo stadio le cellule germinali presentano ancora il complemento cromosomico dello zigote, e si differenzieranno in tessuto ovarico o testicolare a seconda della costituzione genetica del soma. Durante il processo di maturazione della linea germinale viene dapprima eliminato uno dei cromosomi sessuali di origine paterna; le cellule risultanti vanno poi incontro alla divisione meiotica che sarà normale nelle femmine ma anomala nei maschi. Solo il corredo cromosomico materno sarà infatti incluso in spermi vitali; per di più i due cromatidi che portano i caratteri sessuali non si separano, così che gli spermi contengono due cromosomi X.

I casi di inattivazione o di eliminazione selettiva di uno o più cromosomi possono essere considerati un elemento di connessione fra il comportamento dei cromosomi regolari e attivi e quello della categoria cromosomica che comprende i cromosomi accessori o suprannumerari. I cromosomi di questo tipo vengono chiamati anche cromosomi B, in contrapposizione ai componenti del regolare complemento che vengono indicati come cromosomi A. I cromosomi B sono presenti solo in parte degli individui di una popolazione e non sono quindi necessari per la normale vita dell'organismo; inoltre sono presenti con frequenze molto diverse in popolazioni differenti, e raggiungono talvolta un numero molto elevato (anche più di 30). Questo tipo di cromosomi è largamente diffuso in molte specie di piante superiori; la presenza di cromosomi soprannumerari è infatti certa per molte specie appartenenti alle famiglie delle graminace, liliace e composite. Nel regno animale i cromosomi soprannumerari sono diffusi soprattutto tra gli insetti; per esempio, la situazione cromosomica già descritta per i coccidi e per *Sciara* è ulteriormente complicata dalla presenza, nelle sole cellule germinali, di un numero variabile di cromosomi soprannumerari. Cromosomi B sono stati osservati anche fra i platelminti, i molluschi e i vertebrati, fra i quali anche alcuni mammiferi (per esempio alcune specie di marsupiali, il topo di campagna, ecc.). Il numero e la frequenza dei cromosomi B varia molto tra le popolazioni e ciò potrebbe derivare o dall'influenza esercitata dai fattori ambientali di differenti habitat, o dall'influenza che differenti genotipi potrebbero esercitare sulla vitalità degli individui che portano cromosomi accessori.

Generalmente un numero elevato di cromosomi B ha effetti negativi sulla fertilità e il vigore di molte specie vegetali, sebbene un basso numero di cromosomi accessori sia in genere ben tollerato e, in alcuni casi, abbia addirittura effetti positivi.

Nonostante l'apparente svantaggio selettivo collegato a un elevato numero di cromosomi B, questi hanno una spiccata tendenza ad aumentare il proprio numero di generazione in generazione, sia in piante che in animali, attraverso diversi meccanismi, il più diffuso dei quali è il raddoppiamento cromosomico in seguito a non disgiunzione. Nella maggior parte delle specie vegetali esaminate i cromosomi B vanno infatti incontro a fenomeni di non disgiunzione durante la mitosi postmeiotica nella formazione del polline; nella linea germinale femminile generalmente non si verificano fenomeni di non disgiunzione. Variazioni nel numero di cromosomi B si hanno anche nella linea somatica; per esempio, in molte specie vegetali il numero dei cromosomi accessori è costante nella linea vegetativa, ma varia nelle infiorescenze, dove la non disgiunzione somatica porta alla formazione di cellule contenenti un numero di cromosomi B raddoppiato accanto a cellule che ne sono prive.

È quindi chiaro che le cellule somatiche di molte specie, animali e vegetali, possono subire variazioni nel loro assetto cromosomico, anche se le cellule della linea germinale conservano il corredo cromosomico originale; il comportamento cromosomico appare così nel suo insieme molto più complesso e duttile di quanto non venga generalmente ritenuto.

GIOCHI MATEMATICI

di Martin Gardner

La raffinata arte di disporre giocatori, pillole e punti nelle caselle appropriate

Potete dimostrare che un gran numero di cittadini statunitensi ha esattamente lo stesso numero di capelli? Che cos'ha in comune questa domanda con il seguente problema? In un cassetto ci sono 60 calzini identici di colore diverso: 10 paia sono rossi, 10 blu e dieci verdi; le calze sono mescolate fra di loro nel cassetto e la stanza è completamente al buio. Qual è il numero minimo di calze da prendere per essere sicuri che ce ne siano due dello stesso colore?

Si prendano ora in considerazione due esempi meno semplici. Siete in grado di dimostrare che, quando una comune frazione a/b viene espressa in forma decimale, il numero risultante sarà un decimale finito o perodico con un periodo non superiore a b? Siete in grado di dimostrare che, se si dispongono come si vuole 5 punti su un triangolo equilatero di lato 1, almeno due punti non disteranno più di 0,5? (Suggerimento: Dividete il triangolo in quattro triangoli equilateri più piccoli di lato 0,5.)

Ciò che questi e migliaia di altri problemi o giochi matematici hanno in comune è che si possono tutti risolvere facendo riferimento a un vecchio ed efficace principio, il principio dell'incasellamento, che alcuni matematici preferiscono chiamare il principio dei cassetti di Dirichlet dal matematico tedesco del diciannovesimo secolo Peter Gustav Lejeune Dirichelet. Il principio dei cassetti è l'argomento dell'articolo di questo mese che non è stato scritto da me, ma da Ross Honsberger, un matematico dell'Università di Waterloo. Tra le sue opere ricordiamo *Ingenuity in Mathematics*, *Mathematical Gems* e *Mathematical Gems II*. Ha inoltre curato le antologie *Mathematical Morsels* e *Mathematical Plums*. Tutte e cinque le opere sono eccellenti fonti di insoliti problemi che possono essere utilizzati come gustosi passatempi matematici. Tutto ciò che segue (fino al mio commento finale) è stato scritto da Honsberger che a proposito della sua trattazione del principio dell'incasellamento dice: «Può essere utile qualcosa di così semplice?»

Si consideri l'enunciato: «Se la somma di due interi supera 100, almeno uno di essi è maggiore di 50.» Il principio che sta dietro questa semplice asserzione sembra banale, ma non è così. Nella sua formulazione più semplice il principio si può enunciare nel modo seguente: se $n+1$ (o più) oggetti devono essere distribuiti in n scatole, qualche scatola deve contenere almeno due oggetti. Più in generale, se $kn+1$ (o più) oggetti vengono distribuiti in n scatole, una scatola deve contenere almeno $k+1$ oggetti.

Anche nella sua formulazione più generale, questo principio dei cassetti stabilisce semplicemente il fatto ovvio che in un insieme di dati non è possibile che tutti i valori siano al di sotto della media o che tutti i valori siano al di sopra della media. Ciò nonostante, il principio è un concetto matematico di fondamentale importanza e di notevole versatilità. Presenteremo ora sette delle sue più belle applicazioni elementari, cominciando da un semplice esempio geometrico.

1. Le facce del poliedro. Provate a contare gli spigoli che circondano le facce di un poliedro e troverete che ci sono due facce racchiuse dallo stesso numero di spigoli. Per dimostrare che questo è sempre vero, è necessario soltanto immaginare di distribuire le facce in un insieme di scatole numerate 3,4...n, in modo tale che una faccia con r spigoli venga messa nella scatola di numero r. Dato che gli spigoli separano tra di loro le facce, una faccia con il numero massimo di spigoli n è circondata da n facce, il che significa che il poliedro deve avere in totale almeno $n+1$ facce. Sulla base del principio dei cassetti, allora, qualche scatola deve contenere almeno due facce. Con ciò la dimostrazione è completa, infatti è semplice dimostrare che ci sono sempre almeno due differenti coppie di facce con lo stesso numero di spigoli.

2. Dieci interi positivi minori di cento. Ecco un'applicazione del principio dell'incasellamento che sconcerterà i vostri amici. Su un insieme S di 10 interi positivi minori di 100 scelti a piacere si potranno sempre operare due scelte, ciascuna completamente diversa dall'altra, di interi la cui somma è identica. Per esempio, nell'insieme 3, 9, 14, 21, 26, 35, 42, 59, 63, 76 si possono scegliere 14, 63 e 35, 42 la cui somma è in entrambi i casi 77; analogamente si possono selezionare 3, 9, 14 la cui somma è 26, che è nell'insieme.

Per stabilire perché ciò è sempre vero, si osservi che nessun sottoinsieme di 10 elementi di S può dare una somma maggiore dei dieci numeri più grandi compresi tra 1 e 100: 90, 91, ...99. La somma di questi numeri è 945 e quindi i sottoinsiemi di S possono essere raggruppati, sulla base della somma dei loro numeri, in scatole numerate 1, 2, ...945. Dato che ciascun elemento di S o appartiene a un particolare sottoinsieme o non gli appartiene, il numero di sottoinsiemi da classificare (escluso l'insieme vuoto che non ha elementi) è 2^{10}-1, ossia 1023. Per il principio dei cassetti, allora, qualche scatola deve contenere (almeno) due sottoinsiemi diversi A e B. Eliminando i numeri che appartengono sia ad A che a B, si ottengono sue sottoinsiemi distinti A' e B' che danno somme uguali. In realtà, poiché ci sono 78 sottoinsiemi in più rispetto al numero delle scatole, ogni sottoinsieme S dà luogo a dozzine di differenti coppie di sottoinsiemi con somma uguale.

3. Le pillole. Dobbiamo questa applicazione del principio dell'incasellamento a Kenneth R. Rebman, matematico della California State University di Hayward. Un medico che sta sperimentando una nuova medicina dà istruzione a un paziente di prendere 48 pillole per un periodo di trenta giorni. Il paziente è libero di distribuire le pillole come gli pare lungo tale periodo purché ne prenda almeno una al giorno e finisca tutte le 48 pillole nel corso dei trenta giorni. Indipendentemente da come il paziente decide di sistemare le cose, ci sarà un lasso di giorni consecutivi in cui il numero totale di pillole prese sarà 11. In effetti, per ogni valore di k da 1 a 30, con esclusione di 16 17 e 18, è sempre possibile trovare un periodo di giorni consecutivi in cui sono state prese in totale k pillole.

Per dimostrare che un particolare valore di k costituisce un'eccezione alla regola è necessario solo trovare una distribuzione delle pillole per la quale non ci sia nessun lasso di giorni in cui vengano prese k pillole. Così i casi $k=16$, $k=17$ e $k=18$ vengono di colpo eliminati dalla seguente distribuzione in cui si prende una pillola al giorno tranne il sedicesimo giorno in cui si prendono 19 pillole:

$$\underbrace{11\ldots1}_{15}\ 19\ \underbrace{11\ldots1}_{14}$$

Si consideri ora il caso $k=11$. Se p_i sta a indicare il numero totale di pillole prese fino alla fine dell'i-esimo giorno, allora p_{30} è uguale a 48 e i numeri positivi p_1, p_2, ...p_{30} formano una successione strettamente crescente $0<p_1<p_2<\ldots<p_{30}=48$. (Il segno $<$ si legge «minore di» e una successione strettamente crescente è una successione in cui ciascun elemento è maggiore del suo predecessore.) Aggiungendo 11 a ciascuno dei numeri della successione si crea una nuova successione strettamente crescente: $11<p_1+11<p_2+11<\ldots<p_{30}+11=59$.

Ci sono trenta numeri p_i nella prima successione e trenta numeri p_i+11 nella seconda e tutti questi 60 numeri positivi sono minori o uguali a 59. Pertanto, per il principio dei cassetti, due di essi devono essere uguali. Non ci sono, tuttavia, due numeri p_i uguali e nessun p_i+11 risulta uguale all'altro. Quindi qualche p_i deve essere uguale a qualche p_j+11, ossia $p_i=p_j+11$ per alcuni valori di i e j. Quindi p_i meno p_j è uguale a 11, da cui deriva che sono state prese esattamente 11 pillole nei giorni consecutivi $j+1$, $j+2$... i.

Questo ragionamento vale per qualsiasi valore di k fino a 11 compreso e verifica la proprietà per l'intero blocco di valori di k da 1 a 11. È un po' più complicato eliminare i rimanenti casi, ma il principio dei cassetti è ancora lo strumento decisivo. Si considerino subito i casi da $k=31$ a $k=47$. Benché questi valori di k sicuramente ammettano soluzioni in molti casi, la seguente famiglia di distribuzioni mostra che nessuno di essi garantisce una soluzione. Quando n è compreso tra 1 e 17, il valore $k=30+n$ viene eliminato dalla seguente successione:

$$(19-n)\ \underbrace{11\ldots1}_{n+11}\ (n+1)\ \underbrace{11\ldots1}_{17-n}$$

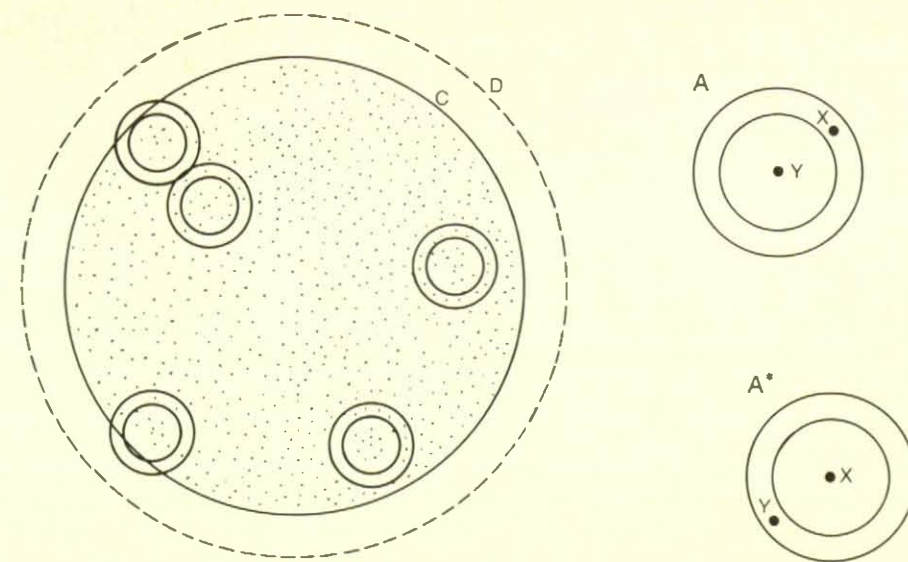
Per esempio, quando n è uguale a 7, la seguente distribuzione elimina il caso $k=37$.

$$12\ \underbrace{11\ldots1}_{18}\ 8\ \underbrace{11\ldots1}_{10}$$

4. 101 numeri. Supponiamo che sui numeri da 1 a 200 venga scelto un insieme di 101 numeri $a_1, a_2, \ldots a_{101}$. Sorprendentemente risulta che è impossibile scegliere tale insieme senza comprendervi due numeri uno dei quali è divisore dell'altro (la divisione cioè non dà resto.) Dimostrare che questa asserzione è vera permette di servirsi di un modo per esprimere gli interi abbastanza chiaro.

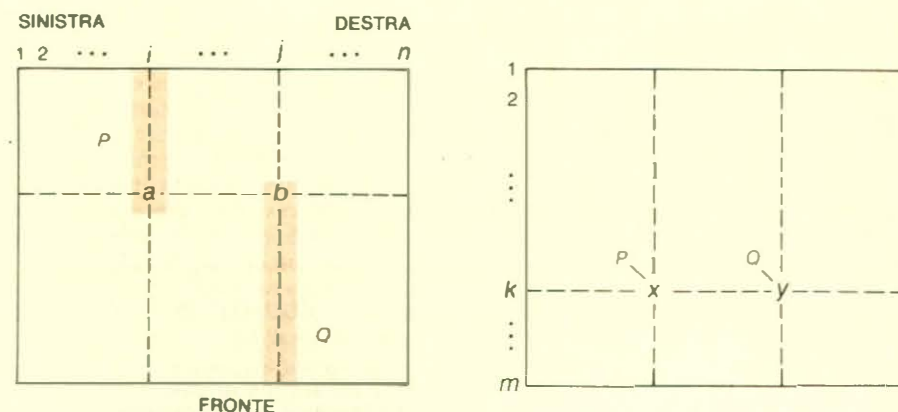
Dato un intero positivo n, è possibile scomporlo in modo che abbia come fattori tanti due quanti può contenere al fine di ridurlo alla forma $n=2^r q$, dove q è un intero dispari (eventualmente uguale a 1). Se ciascuno dei numeri scelti a_i viene espresso in questa forma, si ottiene un insieme di 100 numeri dispari 1, 3, 5, ...199. Sulla base del principio dei cassetti si può concludere che due di questi valori di q devono essere uguali e, quindi, per qualche intero i e j, a_i è uguale a $2^n q$ e a_j è uguale a $2^n q$. Di questi due numeri quello con la più piccola potenza di 2 è divisore dell'altro.

Analogamente non è difficile servirsi del principio dei cassetti per dimostrare che un insieme S di 102 numeri appartenenti all'insieme 1, 2, ...200 deve avere due numeri distinti la cui somma è un terzo numero di S. (Non è qui necessario usare la forma $2^r q$.) Passerò ora a due spettacolari applicazioni del principio dei cassetti a strutture geometriche.

5. Seicentocinquanta punti di un cerchio. Si prenda un cerchio C con un raggio di 16 e un anello A con un raggio esterno di 3 e un raggio interno di 2. Non è degno di nota il fatto che, dovunque si sparga all'interno di C un insieme di 650 punti S, l'anello A può sempre essere disposto sulla figura in modo da coprire almeno 10 punti? Per dimostrare la verità di tale affermazione si potrebbero disporre 650 copie dell'anello A sulla regione circondata dalla circonferenza C in modo tale che ciascun punto di S sia il centro di uno degli anelli, come si può vedere nella figura a sinistra nell'illustrazione in alto. Per i punti di S vicini alla circonferenza, l'anello corrispondente si estenderà al di fuori del cerchio. D'altronde un cerchio concentrico a C con un raggio 19 (pari al raggio di C più il raggio esterno di A) comprenderà sicuramente tutte le copie di A. Si chiami questo nuovo cerchio D. Si noti che l'area di D è $\pi 19^2$, ossia 361π e, dato che l'area di A è $\pi 3^2-\pi 2^2$ ossia 5π, 650 copie di A hanno un'area totale di $3{,}250\pi$.

A questo punto si può applicare una versione «continua» del principio dei cassetti. Ciascuna copia di A copre certe parti del cerchio D quando viene disposta sulla figura. Si supponga che, quando tutte le copie sono state sistemate, non ci sia nessuna parte di D che giaccia sotto più di nove copie differenti di A. In tal caso l'area totale delle copie non potrebbe superare di nove volte l'area di D. Non può tuttavia essere così, perché $9(361\pi)$ è solo $3{,}249\pi$ mentre gli anelli hanno un'area totale di $3{,}250\pi$. Il principio dei cassetti comporta allora che qualche punto X di D deve essere coperto da almeno 10 copie di A.

Supponiamo ora che Y sia un punto di S al centro di una di queste 10 copie di A; allora la distanza tra X e Y deve essere maggiore del raggio interno di A e minore del raggio esterno e, come si vede nella figura in alto a destra, un'altra copia di A con centro in X coprirebbe Y. Chiamiamo questa copia A^*. Poiché ci sono almeno nove altri centri come Y, A^* deve coprire almeno 10 punti di S. L'asserzione è così dimostrata. (Questo problema è stato proposto da Viktors Linis della Università di Ottawa in «Crux Mathematicorum» Vol. 5, pag. 271; novembre 1979).

6. Arriva la banda. L'esempio successivo riguarda una banda i cui membri sono allineati secondo una disposizione

Seicentocinquanta punti in un cerchio.

Dimostrazione del teorema della banda.

rettangolare di m righe e n colonne. Il capobanda, osservando la banda da sinistra, nota che in tale disposizione alcuni membri meno alti rimangono nascosti. Egli, allora, elimina questa pecca estetica disponendo in ogni fila i musicisti in ordine di altezza non decrescente da sinistra a destra in modo che ciascuno è di altezza maggiore o uguale a quella della persona che gli sta a sinistra (dal punto di vista del capobanda). Quando il capobanda torna davanti, tuttavia, trova che ancora una volta non si vedono alcuni dei membri più bassi. Egli procede allora a mescolare i musicisti all'interno delle colonne in modo che siano disposti in ordine di altezza non decrescente. A questo punto egli esita a tornare a sinistra per constatare che cosa questo spostamento ha causato alle righe da lui così accuratamente disposte. Quando, tuttavia, si decide a farlo, è piacevolmente sorpreso nel trovare che anche le righe sono sempre sistemate in ordine di altezza non decrescente da sinistra a destra! Risistemare in questo modo le colonne non modifica l'ordine non decrescente delle righe.

Si può dimostrare questo fatto sorprendente in maniera indiretta ammettendo che sia falso e giungendo a una contraddizione. In altre parole ammetteremo che, dopo che le colonne sono state sistemate, ci sia una riga in cui un musicista più alto a venga messo davanti o a sinistra di uno più basso b. Chiamiamo i la colonna del musicista più alto a e j quella del musicista più basso b, come si può vedere a sinistra nella figura in basso della pagina precedente. Dato che le colonne sono state appena sistemate, si può

Q♥ 5♥ K♦ 9♣ 4♥ 7♣
3♣ 5♦ A♦ 4♦ 6♦ A♠
5♠ 3♠ 3♦ 2♣ 9♠ 8♣
10♦ 10♣ 7♦ J♠ 6♥ K♠

4♥ 5♥ 7♣ 9♣ Q♥ K♦
A♦ A♠ 3♣ 4♦ 5♦ 6♦
2♣ 3♠ 3♦ 5♠ 8♣ 9♠
6♥ 7♦ 10♦ 10♣ J♠ K♠

A♦ A♠ 3♣ 4♦ 5♦ 6♦
2♣ 3♠ 3♦ 5♠ 8♣ 9♠
4♥ 5♥ 7♣ 9♣ J♠ K♦
6♥ 7♦ 10♦ 10♣ Q♥ K♠

Altra versione del teorema della banda.

supporre che ogni musicista del segmento P da a in poi nella colonna i sia almeno altrettanto alto di a e che ogni musicista del segmento Q da b in avanti nella colonna j non sia più alto di b. Inoltre, dato che a è più alto di b, ne segue che ogni membro di P è più alto di ogni membro di Q.

Si consideri ora il punto intermedio in cui sono state sistemate le righe ma non le colonne. Per tornare indietro a questo punto è necessario riportare i musicisti del segmento P alla loro posizione originaria lungo la colonna i e quelli del segmento Q alla loro primitiva posizione nella colonna j. In altre parole i membri di P e Q verrebbero distribuiti sulle righe $1, 2, \ldots, m$ come se le m righe fossero scatole. I segmenti P e Q, tuttavia, hanno una lunghezza totale di $m+1$, cioè c'è un totale di $m+1$ musicisti nei due segmenti. Per il principio dei cassetti due dei musicisti finirebbero nella stessa riga. Essi non potrebbero entrambi venire dallo stesso segmento e così in una certa riga ci dovrebbero essere un x appartenente a P nella colonna i davanti o a sinistra di un y appartenente a Q nella colonna j, come si può vedere a destra nell'illustrazione in basso della pagina precedente. Dato che x è più alto di y, questa disposizione viola l'ordine non decrescente delle righe già stabilito e la conclusione segue per assurdo.

7. Sottosuccessioni in una permutazione. Questo esempio finale porta a stabilire un'interessante proprietà di qualsiasi disposizione di numeri appartenenti alla successione $1, 2, \ldots n^2+1$. Se si esamina ciascuna sistemazione da sinistra a destra, si nota che deve contenere o una sottosuccessione crescente di lunghezza (almeno) $n+1$ o una sottosuccessione decrescente di lunghezza (almeno) $n+1$. Per esempio, quando $n=3$, la disposizione 6, 5, 9, 3, 7, 1, 2, 8, 4, 10 comprende la sottosuccessione decrescente 6, 5, 3, 1. (Come questo esempio dimostra, una sottosuccessione non consiste necessariamente di elementi consecutivi della disposizione.)

L'affermazione che ogni disposizione comprende una sottosuccessione di questo tipo si può dimostrare facilmente precisando che, per ciascun numero i dell'allineamento, x sta per la lunghezza della più lunga sottosuccessione crescente che comincia con i e y sta per la lunghezza della sottosuccessione decrescente che comincia con i.

In questo modo, per l'allineamento di numeri si ottengono n^2+1 coppie di «coordinate» (x, y) e se ogni valore di x o y è pari a $n+1$, l'affermazione è valida. D'altra parte, se ogni valore di x e y è minore o uguale a n, ci sono solo n^2 coppie (x, y) differenti possibili. In questo caso il principio dei cassetti comporta che una coppia (x, y) rappresenti le coordinate di almeno due numeri i e j dell'allineamento. Ma i non è uguale a j e, se i è minore di j, allora la coordinata x di i sarebbe maggiore di quella di j e, se i è maggiore di j, allora la coordinata y di i sarebbe maggiore di quella di j. In ogni caso si arriva a una contraddizione, come si voleva.

Chiudiamo con tre esercizi che speriamo il lettore gradisca e di cui daremo la soluzione il mese prossimo.

1. Un punto reticolare è un punto del piano cartesiano le cui coordinate sono entrambe interi. Si dimostri che, indipendentemente da come si possono scegliere nel piano cinque punti reticolari, almeno uno dei segmenti che uniscono due dei punti scelti deve passare attraverso un punto reticolare del piano.

2. Si dispongano nel piano 6 cerchi (compresa la loro circonferenza) in modo che nessuno di loro contenga il centro di un altro. Si dimostri che non possono avere un punto in comune.

3. Si dimostri che in qualsiasi allineamento di $mn+1$ numeri reali diversi ci deve essere o una sottosuccessione crescente di lunghezza (almeno) $m+1$ o una sottosuccessione decrescente di lunghezza (almeno) $n+1$.

Il risultato controintuitivo che Honsberger descrive a proposito della banda si può efficacemente dimostrare con un mazzo di carte da gioco. Mescolate e distribuite le carte rivoltate secondo una disposizione qualsiasi, diciamo quattro righe di sei carte ciascuna, come si vede in alto nella figura di questa pagina, e poi ridisponete le carte in ogni fila in modo che, da sinistra a destra, i valori non decrescano mai, come si può vedere nella stessa figura al centro. (Per esempio, una disposizione accettabile può essere 6, 7, 10, F, R.) Ora risistemate ogni colonna in modo che dall'alto in basso i valori non decrescano mai, come si può vedere nella solita figura in basso. Permutare le carte in una colonna altera ovviamente la disposizione delle carte nelle righe; ciò nonostante, troverete che, dopo aver ordinato le colonne, anche le righe rimangono ordinate!

Sul periodico di magia «The Pallbearers Review» (pag. 513, aprile 1972) è apparso un gioco con le carte basato su questa sorprendente scoperta. Distribuite le carte come se dovessero andare a cinque giocatori di poker e poi ordinate ciascun gruppetto di cinque carte in modo che i loro valori siano in ordine crescente dal basso in alto. Riunite i mazzetti nel modo che volete e distribuite nuovamente le carte coperte nel modo convenzionale. Le cinque carte di ciascun giocatore saranno completamente differenti e non in ordine. A questo proposito, spiegherete che state cercando di insegnare alle carte a ordinarsi da sole. Prendete quindi ciascun mucchietto e ordinate ancora una volta le carte e deponetele coperte; riunite i gruppetti mettendo il quinto (quello del mazziere) sopra il quarto, questi due sopra il terzo e così via. Distribuite nuovamente le carte al solito modo e ... le carte avranno imparato la lezione: benché le carte di ciascun giocatore siano ancora cambiate, saranno ordinate!

Questo risultato fa parte della teoria dei tableaux di Young, una classe di distribuzioni di numeri in righe e colonne che prende il nome dal reverendo inglese Alfred Young che le propose e le analizzò in un articolo pubblicato nel 1900. Si è

dimostrato che tali allineamenti hanno importanti applicazioni nella meccanica quantistica.

Ecco le risposte ai problemi relativi ai giochi lineari del mese scorso.

Nella dama lineare a 11 caselle (cominciando con le pedine nere sulle caselle 1, 2, 3 e 4 e le pedine bianche sulle caselle 8, 9, 10 e 11) le prime due mosse sono obbligate: il Nero su 5 e il Bianco su 7. Per evitare di perdere, il Nero va poi su 4 e il Bianco deve rispondere muovendo su 8. Il Nero è allora obbligato ad andare su 3 e il Bianco su 9. A questo punto il Bianco perde se muove su 2, ma vince se muove su 6. In quest'ultimo caso il Bianco salta su 5 e il Nero salta su 6 ottenendo una facile vittoria.

Sulla scacchiera lineare a 8 caselle il Bianco può vincere in un massimo di 6 mosse. Delle quattro mosse di apertura del Bianco, T×T porta a uno stallo istantaneo e costituisce il gioco più breve possibile. T-5 porta rapidamente il Bianco alla sconfitta se il Nero muove T×T. In questo caso il Bianco deve rispondere con C-4 e il Nero dà scacco matto alla seconda mossa con T×C. Questo è uno dei più brevi giochi possibili che portano alla vittoria. L'apertura T-4 permette al Nero di dare scacco alla seconda o alla terza mossa se risponde con C-5.

La sola apertura vincente per il Bianco è C-4. In questo caso il Nero ha tre possibili risposte:

1. T×C. In questo caso il Bianco vince in due mosse con T×T.
2. T-5. Il Bianco vince con R-2. Se il Nero gioca T-6, il Bianco dà scacco con T×C. Se il Nero prende il Cavallo, il Bianco prende la Torre, il Nero muove C-5 e il Bianco dà scacco matto prendendo il cavallo nero.
3. C-5 Questa mossa ritarda la sconfitta del Nero e rende il gioco il più lungo dei tre. Per vincere, il Bianco deve dare scacco con C×T, costringendo il Re nero in 7. Il Bianco muove la sua Torre in 4. Se il Nero muove R×C, il Re bianco va in 2, il Re nero è obbligato in 7 e il Bianco muove T×C, vincendo. Se il Nero muove C-3 (scacco), il Bianco sposta il Re in 2. Il Nero può muovere solo il cavallo. Se lo mette in 1, il Bianco dà scacco matto con C-8. Se il Nero muove C-5, il Bianco muove C-8 obbligando il Nero a R×C e quindi il Bianco dà scacco matto con T×C.

Anche nel gioco lineare a otto caselle il primo giocatore vince, muovendo in apertura sulla seconda casella a partire da un estremo, mossa che porta alla vittoria anche nel gioco con sei e sette caselle. Ammettiamo che il primo giocatore vada sulla casella due. Le sue uniche risposte vincenti al gioco dell'avversario in 3, 4, 5, 6, 7 e 8 sono rispettivamente 5, 7, 7, 7, 5 e 6. Lascio il resto del gioco al lettore. Non si sa se ci siano altre mosse di apertura vincenti. James Henel, l'inventore del gioco, mi assicura che nel gioco con nove caselle la vittoria spetta al secondo giocatore, mentre non ha provato ad analizzare scacchiere con più di nove caselle.